Bigdata Interview Preparation Guide

(1000+ Questions with Answers)

Programming, Scenario-Based, Fundamentals, & Performance Tunning

WWW.SMARTDATACAMP.COM

Table Of Content Page

www.smartdatacamp.com

Apache Hadoop

Apache Hadoop is an open-source software framework used for distributed storage and processing of dataset of big data using the MapReduce programming model

1)How does Hadoop Namenode failover process works?

Answer)In a typical High Availability cluster, two separate machines are configured as NameNodes. At any point in time, exactly one of the NameNodes is in an Active state, and the other is in a Standby state. The Active NameNode is responsible for all client operations in the cluster, while the Standby is simply acting as a slave, maintaining enough state to provide a fast failover if necessary.

In order for the Standby node to keep its state synchronized with the Active node, both nodes communicate with a group of separate daemons called "JournalNodes" (JNs). When any namespace modification is performed by the Active node, it durably logs a record of the modification to a majority of these JNs. The Standby node is capable of reading the edits from the JNs, and is constantly watching them for changes to the edit log. As the Standby Node sees the edits, it applies them to its own namespace. In the event of a failover, the Standby will ensure that it has read all of the edits from the JounalNodes before promoting itself to the Active state. This ensures that the namespace state is fully synchronized before a failover occurs.

In order to provide a fast failover, it is also necessary that the Standby node have up-to-date information regarding the location of blocks in the cluster. In order to achieve this, the DataNodes are configured with the location of both NameNodes, and send block location information and heartbeats to both.

It is vital for the correct operation of an HA cluster that only one of the NameNodes be Active at a time. Otherwise, the namespace state would quickly diverge between the two, risking data loss or other incorrect results. In order to ensure this property and prevent the so-called "split-brain scenario," the JournalNodes will only ever allow a single NameNode to be a writer at a time. During a failover, the NameNode which is to become active will simply take over the role of writing to the JournalNodes, which will effectively prevent the other NameNode from continuing in the Active state, allowing the new Active to safely proceed with failover.

2) How can we initiate a manual failover when automatic failover is configured?

Answer: Even if automatic failover is configured, you may initiate a manual failover using the same hdfs haadmin command. It will perform a coordinated failover.

3) When not use Hadoop?

www.smartdatacamp.com

Answer: 1)Real Time Analytics: If you want to do some Real Time Analytics, where you are expecting result quickly, Hadoop should not be used directly. It is because Hadoop works on batch processing, hence response time is high.

2. Not a Replacement for Existing Infrastructure: Hadoop is not a replacement for your existing data processing infrastructure. However, you can use Hadoop along with it.

3. Multiple Smaller Datasets:Hadoop framework is not recommended for small-structured datasets as you have other tools available in market which can do this work quite easily and at a fast pace than Hadoop like MS Excel, RDBMS etc. For a small data analytics, Hadoop can be costlier than other tools.

4. Novice Hadoopers:Unless you have a better understanding of the Hadoop framework, it's not suggested to use Hadoop for production. Hadoop is a technology which should come with a disclaimer: "Handle with care". You should know it before you use it or else you will end up like the kid below.

5. Security is the primary Concern:Many enterprises especially within highly regulated industries dealing with sensitive data aren't able to move as quickly as they would like towards implementing Big Data projects and Hadoop.

4) When To Use Hadoop?

Answer: 1. Data Size and Data Diversity:When you are dealing with huge volumes of data coming from various sources and in a variety of formats then you can say that you are dealing with Big Data. In this case, Hadoop is the right technology for you.

2. Future Planning: It is all about getting ready for challenges you may face in future. If you anticipate Hadoop as a future need then you should plan accordingly. To implement Hadoop on you data you should first understand the level of complexity of data and the rate with which it is going to grow. So, you need a cluster planning. It may begin with building a small or medium cluster in your industry as per data (in GBs or few TBs) available at present and scale up your cluster in future depending on the growth of your data.

3. Multiple Frameworks for Big Data: There are various tools for various purposes. Hadoop can be integrated with multiple analytic tools to get the best out of it, like Mahout for Machine-Learning, R and Python for Analytics and visualization, Python, Spark for real time processing, MongoDB and Hbase for Nosql database, Pentaho for BI etc.

4. Lifetime Data Availability: When you want your data to be live and running forever, it can be achieved using Hadoop's scalability. There is no limit to the size of cluster that you can have. You can increase the size anytime as per your need by adding datanodes to it with The bottom line is use the right technology as per your need.

5) When you run start-dfs.sh or stop-dfs.sh, you get the following warning message:WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable.How to fix this warning message?

Answer) The reason you saw that warning is the native Hadoop library $HADOOP_HOME/lib/native/libhadoop.so.1.0.0 was actually compiled on 32 bit.Anyway, it's just a warning, and won't impact Hadoop's functionalities. Here is the way if you do want to eliminate this warning, download the source code of Hadoop and recompile libhadoop.so.1.0.0 on 64bit system, then replace the 32bit one.

6) What platforms and Java versions does Hadoop run on?

Answer)Java 1.6.x or higher, preferably, Linux and Windows are the supported operating systems, but BSD, Mac OS/X, and OpenSolaris are known to work. (Windows requires the installation of Cygwin).

7) As we talk about Hadoop is Highly scalable how well does it Scale?

Answer) Hadoop has been demonstrated on clusters of up to 4000 nodes. Sort performance on 900 nodes is good (sorting 9TB of data on 900 nodes takes around 1.8 hours) and improving using these non-default configuration values:
dfs.block.size = 134217728
dfs.namenode.handler.count = 40
mapred.reduce.parallel.copies = 20
mapred.child.java.opts = -Xmx512m
fs.inmemory.size.mb = 200
io.sort.factor = 100
io.sort.mb = 200
io.file.buffer.size = 131072

Sort performances on 1400 nodes and 2000 nodes are pretty good too - sorting 14TB of data on a 1400-node cluster takes 2.2 hours; sorting 20TB on a 2000-node cluster takes 2.5 hours. The updates to the above configuration being:
mapred.job.tracker.handler.count = 60
mapred.reduce.parallel.copies = 50
tasktracker.http.threads = 50
mapred.child.java.opts = -Xmx1024m

8) What kind of hardware scales best for Hadoop?

Answer) The short answer is dual processor/dual core machines with 4-8GB of RAM using ECC memory, depending upon workflow needs. Machines should be moderately high-end commodity machines to be most cost-effective and typically cost 1/2 - 2/3 the cost of normal production application servers but are not desktop-class machines.

www.smartdatacamp.com

9) Among the software questions for setting up and running Hadoop, there a few other questions that relate to hardware scaling:

i)What are the optimum machine configurations for running a hadoop cluster?
ii) Should I use a smaller number of high end/performance machines or are a larger number of "commodity" machines?
iii)How does the Hadoop/Parallel Distributed Processing community define "commodity"?

Answer)In answer to i and ii above, we can group the possible hardware options in to 3 rough categories:
A)Database class machine with many (>10) fast SAS drives and >10GB memory, dual or quad x quad core cpu's. With an approximate cost of $20K.
B)Generic production machine with 2 x 250GB SATA drives, 4-12GB RAM, dual x dual core CPU's (=Dell 1950). Cost is about $2-5K.
C) POS beige box machine with 2 x SATA drives of variable size, 4 GB RAM, single dual core CPU. Cost is about $1K.

For a $50K budget, most users would take 25xB over 50xC due to simpler and smaller admin issues even though cost/performance would be nominally about the same. Most users would avoid 2x(A) like the plague.

For the discussion to iii, "commodity" hardware is best defined as consisting of standardized, easily available components which can be purchased from multiple distributors/retailers. Given this definition there are still ranges of quality that can be purchased for your cluster. As mentioned above, users generally avoid the low-end, cheap solutions. The primary motivating force to avoid low-end solutions is "real" cost; cheap parts mean greater number of failures requiring more maintanance/cost. Many users spend $2K-$5K per machine.

More specifics:

Multi-core boxes tend to give more computation per dollar, per watt and per unit of operational maintenance. But the highest clockrate processors tend to not be cost-effective, as do the very largest drives. So moderately high-end commodity hardware is the most cost-effective for Hadoop today.

Some users use cast-off machines that were not reliable enough for other applications. These machines originally cost about 2/3 what normal production boxes cost and achieve almost exactly 1/2 as much. Production boxes are typically dual CPU's with dual cores.

RAM:

Many users find that most hadoop applications are very small in memory consumption. Users tend to have 4-8 GB machines with 2GB probably being too little. Hadoop benefits greatly from ECC memory, which is not low-end, however using ECC memory is RECOMMENDED

10) I have a new node I want to add to a running Hadoop cluster; how do I start services on just one node?

Answer)This also applies to the case where a machine has crashed and rebooted, etc, and you need to get it to rejoin the cluster. You do not need to shutdown and/or restart the entire cluster in this case.

First, add the new node's DNS name to the conf/slaves file on the master node.

Then log in to the new slave node and execute:
$ cd path/to/hadoop
$ bin/hadoop-daemon.sh start datanode
$ bin/hadoop-daemon.sh start tasktracker

If you are using the dfs.include/mapred.include functionality, you will need to additionally add the node to the dfs.include/mapred.include file, then issue hadoop dfsadmin -refreshNodes and hadoop mradmin -refreshNodes so that the NameNode and JobTracker know of the additional node that has been added.

11) Is there an easy way to see the status and health of a cluster?

Answer) You can also see some basic HDFS cluster health data by running:

$ bin/hadoop dfsadmin -report

12) How much network bandwidth might I need between racks in a medium size (40-80 node) Hadoop cluster?

Answer) The true answer depends on the types of jobs you're running. As a back of the envelope calculation one might figure something like this:

60 nodes total on 2 racks = 30 nodes per rack Each node might process about 100MB/sec of data In the case of a sort job where the intermediate data is the same size as the input data, that means each node needs to shuffle 100MB/sec of data In aggregate, each rack is then producing about 3GB/sec of data However, given even reducer spread across the racks, each rack will need to send 1.5GB/sec to reducers running on the other rack. Since the connection is full duplex, that means you need 1.5GB/sec of bisection bandwidth for this theoretical job. So that's 12Gbps.

However, the above calculations are probably somewhat of an upper bound. A large number of jobs have significant data reduction during the map phase, either by some kind of filtering/selection going on in the Mapper itself, or by good usage of Combiners. Additionally, intermediate data compression can cut the intermediate data transfer by a significant factor. Lastly, although your disks can probably provide 100MB sustained throughput, it's rare to see a MR job which can sustain disk speed IO through the entire pipeline. So, I'd say my estimate is at least a factor of 2 too high.

www.smartdatacamp.com

So, the simple answer is that 4-6Gbps is most likely just fine for most practical jobs. If you want to be extra safe, many inexpensive switches can operate in a "stacked" configuration where the bandwidth between them is essentially backplane speed. That should scale you to 96 nodes with plenty of headroom. Many inexpensive gigabit switches also have one or two 10GigE ports which can be used effectively to connect to each other or to a 10GE core.

13) I am seeing connection refused in the logs. How do I troubleshoot this?

Answer) You get a ConnectionRefused Exception when there is a machine at the address specified, but there is no program listening on the specific TCP port the client is using -and there is no firewall in the way silently dropping TCP connection requests.

Unless there is a configuration error at either end, a common cause for this is the Hadoop service isn't running.

This stack trace is very common when the cluster is being shut down -because at that point Hadoop services are being torn down across the cluster, which is visible to those services and applications which haven't been shut down themselves. Seeing this error message during cluster shutdown is not anything to worry about.

If the application or cluster is not working, and this message appears in the log, then it is more serious.

Check the hostname the client using is correct. If it's in a Hadoop configuration option: examine it carefully, try doing an ping by hand.

Check the IP address the client is trying to talk to for the hostname is correct.

Make sure the destination address in the exception isn't 0.0.0.0 -this means that you haven't actually configured the client with the real address for that service, and instead it is picking up the server-side property telling it to listen on every port for connections.

If the error message says the remote service is on "127.0.0.1" or "localhost" that means the configuration file is telling the client that the service is on the local server. If your client is trying to talk to a remote system, then your configuration is broken.

Check that there isn't an entry for your hostname mapped to 127.0.0.1 or 127.0.1.1 in /etc/hosts (Ubuntu is notorious for this).

Check the port the client is trying to talk to using matches that the server is offering a service on. The netstat command is useful there.

On the server, try a telnet localhost (port) to see if the port is open there.

On the client, try a telnet (server) (port) to see if the port is accessible remotely.

Try connecting to the server/port from a different machine, to see if it just the single client misbehaving.

If your client and the server are in different subdomains, it may be that the configuration of the service is only publishing the basic hostname, rather than the Fully Qualified Domain Name. The client in the different subdomain can be unintentionally attempt to bind to a host in the local subdomain —and failing.

If you are using a Hadoop-based product from a third party, -please use the support channels provided by the vendor.

14)Does Hadoop require SSH?

Answer) Hadoop provided scripts (e.g., start-mapred.sh and start-dfs.sh) use ssh in order to start and stop the various daemons and some other utilities. The Hadoop framework in itself does not require ssh. Daemons (e.g. TaskTracker and DataNode) can also be started manually on each node without the script's help.

15) What does NFS: Cannot create lock on (some dir) mean?

Answer) This actually is not a problem with Hadoop, but represents a problem with the setup of the environment it is operating.

Usually, this error means that the NFS server to which the process is writing does not support file system locks. NFS prior to v4 requires a locking service daemon to run (typically rpc.lockd) in order to provide this functionality. NFSv4 has file system locks built into the protocol.

In some (rarer) instances, it might represent a problem with certain Linux kernels that did not implement the flock() system call properly.

It is highly recommended that the only NFS connection in a Hadoop setup be the place where the NameNode writes a secondary or tertiary copy of the fsimage and edits log. All other users of NFS are not recommended for optimal performance.

16) If I add new DataNodes to the cluster will HDFS move the blocks to the newly added nodes in order to balance disk space utilization between the nodes?

Answer) No, HDFS will not move blocks to new nodes automatically. However, newly created files will likely have their blocks placed on the new nodes.

There are several ways to rebalance the cluster manually.

Select a subset of files that take up a good percentage of your disk space; copy them to new locations in HDFS; remove the old copies of the files; rename the new copies to their original names.

A simpler way, with no interruption of service, is to turn up the replication of files, wait for transfers to stabilize, and then turn the replication back down.

Yet another way to re-balance blocks is to turn off the data-node, which is full, wait until its blocks are replicated, and then bring it back again. The over-replicated blocks will be randomly removed from different nodes, so you really get them rebalanced not just removed from the current node.

Finally, you can use the bin/start-balancer.sh command to run a balancing process to move blocks around the cluster automatically.

www.smartdatacamp.com

17) What is the purpose of the secondary name-node?

The term "secondary name-node" is somewhat misleading. It is not a name-node in the sense that data-nodes cannot connect to the secondary name-node, and in no event it can replace the primary name-node in case of its failure.

The only purpose of the secondary name-node is to perform periodic checkpoints. The secondary name-node periodically downloads current name-node image and edits log files, joins them into new image and uploads the new image back to the (primary and the only) name-node.

So if the name-node fails and you can restart it on the same physical node then there is no need to shutdown data-nodes, just the name-node need to be restarted. If you cannot use the old node anymore you will need to copy the latest image somewhere else. The latest image can be found either on the node that used to be the primary before failure if available; or on the secondary name-node. The latter will be the latest checkpoint without subsequent edits logs, that is the most recent name space modifications may be missing there. You will also need to restart the whole cluster in this case.

18) Does the name-node stay in safe mode till all under-replicated files are fully replicated?

Answer) No. During safe mode replication of blocks is prohibited. The name-node awaits when all or majority of data-nodes report their blocks.

Depending on how safe mode parameters are configured the name-node will stay in safe mode until a specific percentage of blocks of the system is minimally replicated dfs.replication.min. If the safe mode threshold dfs.safemode.threshold.pct is set to 1 then all blocks of all files should be minimally replicated.

Minimal replication does not mean full replication. Some replicas may be missing and in order to replicate them the name-node needs to leave safe mode.

19) How do I set up a hadoop node to use multiple volumes?

Answer) Data-nodes can store blocks in multiple directories typically allocated on different local disk drives. In order to setup multiple directories one needs to specify a comma separated list of pathnames as a value of the configuration parameter dfs.datanode.data.dir. Data-nodes will attempt to place equal amount of data in each of the directories.

The name-node also supports multiple directories, which in the case store the name space image and the edits log. The directories are specified via the dfs.namenode.name.dir configuration parameter. The name-node directories are used for the name space data replication so that the image and the log could be restored from the remaining volumes if one of them fails.

20) What happens if one Hadoop client renames a file or a directory containing this file while another client is still writing into it?

Answer)Starting with release hadoop-0.15, a file will appear in the name space as soon as it is created. If a writer is writing to a file and another client renames either the file itself or any of its path components, then the original writer will get an IOException either when it finishes writing to the current block or when it closes the file.

21)I want to make a large cluster smaller by taking out a bunch of nodes simultaneously. How can this be done?

Answer) On a large cluster removing one or two data-nodes will not lead to any data loss, because name-node will replicate their blocks as long as it will detect that the nodes are dead. With a large number of nodes getting removed or dying the probability of losing data is higher.

Hadoop offers the decommission feature to retire a set of existing data-nodes. The nodes to be retired should be included into the exclude file, and the exclude file name should be specified as a configuration parameter dfs.hosts.exclude. This file should have been specified during namenode startup. It could be a zero length file. You must use the full hostname, ip or ip:port format in this file. (Note that some users have trouble using the host name. If your namenode shows some nodes in "Live" and "Dead" but not decommission, try using the full ip:port.) Then the shell command

bin/hadoop dfsadmin -refreshNodes

should be called, which forces the name-node to re-read the exclude file and start the decommission process.

Decommission is not instant since it requires replication of potentially a large number of blocks and we do not want the cluster to be overwhelmed with just this one job. The decommission progress can be monitored on the name-node Web UI. Until all blocks are replicated the node will be in "Decommission In Progress" state. When decommission is done the state will change to "Decommissioned". The nodes can be removed whenever decommission is finished.

The decommission process can be terminated at any time by editing the configuration or the exclude files and repeating the -refreshNodes command.

22) Does Wildcard characters work correctly in FsShell?

Answer)When you issue a command in FsShell, you may want to apply that command to more than one file. FsShell provides a wildcard character to help you do so. The * (asterisk) character can be used to take the place of any set of characters. For example, if you would like to list all the files in your account which begin with the letter x, you could use the ls command with the * wildcard:

bin/hadoop dfs -ls x*

Sometimes, the native OS wildcard support causes unexpected results. To avoid this problem, Enclose the expression in Single or Double quotes and it should work correctly.

www.smartdatacamp.com

bin/hadoop dfs -ls 'in*'

23) Can I have multiple files in HDFS use different block sizes?

Answer)Yes. HDFS provides api to specify block size when you create a file.

See FileSystem.create(Path, overwrite, bufferSize, replication, blockSize, progress)

24) Does HDFS make block boundaries between records?

Answer) No, HDFS does not provide record-oriented API and therefore is not aware of records and boundaries between them.

25) What happens when two clients try to write into the same HDFS file?

Answer)HDFS supports exclusive writes only.

When the first client contacts the name-node to open the file for writing, the name-node grants a lease to the client to create this file. When the second client tries to open the same file for writing, the name-node will see that the lease for the file is already granted to another client, and will reject the open request for the second client.

26) How to limit Data node's disk usage?

Answer) Use dfs.datanode.du.reserved configuration value in $HADOOP_HOME/conf/hdfs-site.xml for limiting disk usage.

value = 182400

27) On an individual data node, how do you balance the blocks on the disk?

Answer) Hadoop currently does not have a method by which to do this automatically. To do this manually:

1) Shutdown the DataNode involved

2) Use the UNIX mv command to move the individual block replica and meta pairs from one directory to another on the selected host. On releases which have HDFS-6482 (Apache Hadoop 2.6.0+) you also need to ensure the subdir-named directory structure remains exactly the same when moving the blocks across the disks. For example, if the block replica and its meta pair were under

/data/1/dfs/dn/current/BP-1788246909-172.23.1.202-1412278461680/current/finalized/subdir0/subdir1/, and you wanted to move it to /data/5/ disk, then it MUST be moved into the same subdirectory structure underneath that, i.e.

/data/5/dfs/dn/current/BP-1788246909-172.23.1.202-1412278461680/current/finalized/subdir0/subdir1/. If this is not maintained, the DN will no longer be able to locate the replicas after the move.

3) Restart the DataNode.

28) What does "file could only be replicated to 0 nodes, instead of 1" mean?

Answer) The NameNode does not have any available DataNodes. This can be caused by a wide variety of reasons. Check the DataNode logs, the NameNode logs, network connectivity.

29) If the NameNode loses its only copy of the fsimage file, can the file system be recovered from the DataNodes?

Answer) No. This is why it is very important to configure dfs.namenode.name.dir to write to two filesystems on different physical hosts, use the SecondaryNameNode, etc.

30) I got a warning on the NameNode web UI "WARNING : There are about 32 missing blocks. Please check the log or run fsck." What does it mean?

Answer) This means that 32 blocks in your HDFS installation don’t have a single replica on any of the live DataNodes.

Block replica files can be found on a DataNode in storage directories specified by configuration parameter dfs.datanode.data.dir. If the parameter is not set in the DataNode’s hdfs-site.xml, then the default location /tmp will be used. This default is intended to be used only for testing. In a production system this is an easy way to lose actual data, as local OS may enforce recycling policies on /tmp. Thus the parameter must be overridden.

If dfs.datanode.data.dir correctly specifies storage directories on all DataNodes, then you might have a real data loss, which can be a result of faulty hardware or software bugs. If the file(s) containing missing blocks represent transient data or can be recovered from an external source, then the easiest way is to remove (and potentially restore) them. Run fsck in order to determine which files have missing blocks. If you would like (highly appreciated) to further investigate the cause of data loss, then you can dig into NameNode and DataNode logs. From the logs one can track the entire life cycle of a particular block and its replicas.

31) If a block size of 64MB is used and a file is written that uses less than 64MB, will 64MB of disk space be consumed?

www.smartdatacamp.com

Answer) Short answer: No.

Longer answer: Since HFDS does not do raw disk block storage, there are two block sizes in use when writing a file in HDFS: the HDFS blocks size and the underlying file system's block size. HDFS will create files up to the size of the HDFS block size as well as a meta file that contains CRC32 checksums for that block. The underlying file system store that file as increments of its block size on the actual raw disk, just as it would any other file.

32) What does the message "Operation category READ/WRITE is not supported in state standby" mean?

Answer) In an HA-enabled cluster, DFS clients cannot know in advance which namenode is active at a given time. So when a client contacts a namenode and it happens to be the standby, the READ or WRITE operation will be refused and this message is logged. The client will then automatically contact the other namenode and try the operation again. As long as there is one active and one standby namenode in the cluster, this message can be safely ignored.

33)On what concept the Hadoop framework works?

Answer) Hadoop Framework works on the following two core components-

1)HDFS – Hadoop Distributed File System is the java based file system for scalable and reliable storage of large datasets. Data in HDFS is stored in the form of blocks and it operates on the Master Slave Architecture.

2)Hadoop MapReduce-This is a java based programming paradigm of Hadoop framework that provides scalability across various Hadoop clusters. MapReduce distributes the workload into various tasks that can run in parallel. Hadoop jobs perform 2 separate tasks- job. The map job breaks down the data sets into key-value pairs or tuples. The reduce job then takes the output of the map job and combines the data tuples to into smaller set of tuples. The reduce job is always performed after the map job is executed.

34)What is Hadoop streaming?

Answer)Hadoop distribution has a generic application programming interface for writing Map and Reduce jobs in any desired programming language like Python, Perl, Ruby, etc. This is referred to as Hadoop Streaming. Users can create and run jobs with any kind of shell scripts or executable as the Mapper or Reducers.

35)Explain about the process of inter cluster data copying.?

Answer)HDFS provides a distributed data copying facility through the DistCP from source to destination. If this data copying is within the hadoop cluster then it is referred to as inter cluster data copying. DistCP requires both source and destination to have a compatible or same version of hadoop.

36)Differentiate between Structured and Unstructured data?

Answer)Data which can be stored in traditional database systems in the form of rows and columns, for example the online purchase transactions can be referred to as Structured Data. Data which can be stored only partially in traditional database systems, for example, data in XML records can be referred to as semi structured data. Unorganized and raw data that cannot be categorized as semi structured or structured data is referred to as unstructured data. Facebook updates, Tweets on Twitter, Reviews, web logs, etc. are all examples of unstructured data.

37)Explain the difference between NameNode, Backup Node and Checkpoint NameNode?

Answer)NameNode: NameNode is at the heart of the HDFS file system which manages the metadata i.e. the data of the files is not stored on the NameNode but rather it has the directory tree of all the files present in the HDFS file system on a hadoop cluster. NameNode uses two files for the namespace-
fsimage file- It keeps track of the latest checkpoint of the namespace.
edits file-It is a log of changes that have been made to the namespace since checkpoint.

Checkpoint Node:

Checkpoint Node keeps track of the latest checkpoint in a directory that has same structure as that of NameNode's directory. Checkpoint node creates checkpoints for the namespace at regular intervals by downloading the edits and fsimage file from the NameNode and merging it locally. The new image is then again updated back to the active NameNode.

BackupNode:

Backup Node also provides check pointing functionality like that of the checkpoint node but it also maintains its up-to-date in-memory copy of the file system namespace that is in sync with the active NameNode.

38)How can you overwrite the replication factors in HDFS?

Answer)The replication factor in HDFS can be modified or overwritten in 2 ways-

1)Using the Hadoop FS Shell, replication factor can be changed per file basis using the below command-

$hadoop fs –setrep –w 2 /my/test_file (test_file is the filename whose replication factor will be set to 2)

www.smartdatacamp.com

2)Using the Hadoop FS Shell, replication factor of all files under a given directory can be modified using the below command-

3)$hadoop fs –setrep –w 5 /my/test_dir (test_dir is the name of the directory and all the files in this directory will have a replication factor set to 5)

39)Explain what happens if during the PUT operation, HDFS block is assigned a replication factor 1 instead of the default value 3?

Answer)Replication factor is a property of HDFS that can be set accordingly for the entire cluster to adjust the number of times the blocks are to be replicated to ensure high data availability. For every block that is stored in HDFS, the cluster will have n-1 duplicated blocks. So, if the replication factor during the PUT operation is set to 1 instead of the default value 3, then it will have a single copy of data. Under these circumstances when the replication factor is set to 1 ,if the DataNode crashes under any circumstances, then only single copy of the data would be lost.

40)What is the process to change the files at arbitrary locations in HDFS?

Answer)HDFS does not support modifications at arbitrary offsets in the file or multiple writers but files are written by a single writer in append only format i.e. writes to a file in HDFS are always made at the end of the file.

41)Explain about the indexing process in HDFS?

Answer)Indexing process in HDFS depends on the block size. HDFS stores the last part of the data that further points to the address where the next part of data chunk is stored.

42)What is a rack awareness and on what basis is data stored in a rack?

Answer)All the data nodes put together form a storage area i.e. the physical location of the data nodes is referred to as Rack in HDFS. The rack information i.e. the rack id of each data node is acquired by the NameNode. The process of selecting closer data nodes depending on the rack information is known as Rack Awareness.

The contents present in the file are divided into data block as soon as the client is ready to load the file into the hadoop cluster. After consulting with the NameNode, client allocates 3 data nodes for each data block. For each data block, there exists 2 copies in one rack and the third copy is present in another rack. This is generally referred to as the Replica Placement Policy.

43)What happens to a NameNode that has no data?

Answer)There does not exist any NameNode without data. If it is a NameNode then it should have some sort of data in it.

44) What happens when a user submits a Hadoop job when the NameNode is down- does the job get in to hold or does it fail.

Answer) The Hadoop job fails when the NameNode is down.

45) What happens when a user submits a Hadoop job when the Job Tracker is down- does the job get in to hold or does it fail.

Answer)The Hadoop job fails when the Job Tracker is down.

46) Whenever a client submits a hadoop job, who receives it?

Answer)NameNode receives the Hadoop job which then looks for the data requested by the client and provides the block information. JobTracker takes care of resource allocation of the hadoop job to ensure timely completion.

47) What do you understand by edge nodes in Hadoop?

Edges nodes are the interface between hadoop cluster and the external network. Edge nodes are used for running cluster adminstration tools and client applications.Edge nodes are also referred to as gateway nodes.

48)What are real-time industry applications of Hadoop?

Answer)Hadoop, well known as Apache Hadoop, is an open-source software platform for scalable and distributed computing of large volumes of data. It provides rapid, high performance and cost-effective analysis of structured and unstructured data generated on digital platforms and within the enterprise. It is used in almost all departments and sectors today.Some of the instances where Hadoop is used:
Managing traffic on streets.
Streaming processing.
Content Management and Archiving Emails.
Processing Rat Brain Neuronal Signals using a Hadoop Computing Cluster.
Fraud detection and Prevention.

www.smartdatacamp.com

Advertisements Targeting Platforms are using Hadoop to capture and analyze click stream, transaction, video and social media data.
Managing content, posts, images and videos on social media platforms.
Analyzing customer data in real-time for improving business performance.
Public sector fields such as intelligence, defense, cyber security and scientific research.
Financial agencies are using Big Data Hadoop to reduce risk, analyze fraud patterns, identify rogue traders, more precisely target their marketing campaigns based on customer segmentation, and improve customer satisfaction.
Getting access to unstructured data like output from medical devices, doctor's notes, lab results, imaging reports, medical correspondence, clinical data, and financial data.

49)What all modes Hadoop can be run in?

Answer)Hadoop can run in three modes:

Standalone Mode: Default mode of Hadoop, it uses local file stystem for input and output operations. This mode is mainly used for debugging purpose, and it does not support the use of HDFS. Further, in this mode, there is no custom configuration required for mapred-site.xml, core-site.xml, hdfs-site.xml files. Much faster when compared to other modes.

Pseudo-Distributed Mode (Single Node Cluster): In this case, you need configuration for all the three files mentioned above. In this case, all daemons are running on one node and thus, both Master and Slave node are the same.

Fully Distributed Mode (Multiple Cluster Node): This is the production phase of Hadoop (what Hadoop is known for) where data is used and distributed across several nodes on a Hadoop cluster. Separate nodes are allotted as Master and Slave.

50)Explain the major difference between HDFS block and InputSplit?

Answer)In simple terms, block is the physical representation of data while split is the logical representation of data present in the block. Split acts as an intermediary between block and mapper.

Suppose we have two blocks:

Block 1: ii bbhhaavveesshhll

Block 2: Ii inntteerrvviieewwll

Now, considering the map, it will read first block from ii till ll, but does not know how to process the second block at the same time. Here comes Split into play, which will form a logical group of Block1 and Block 2 as a single block.

It then forms key-value pair using inputformat and records reader and sends map for further processing With inputsplit, if you have limited resources, you can increase the split size to limit the number of maps. For instance, if there are 10 blocks of 640MB (64MB each) and there are limited resources, you can assign 'split size' as 128MB. This will form a logical group of 128MB, with only 5 maps executing at a time.

However, if the 'split size' property is set to false, whole file will form one inputsplit and is processed by single map, consuming more time when the file is bigger.

51)What are the most common Input Formats in Hadoop?

Answer)There are three most common input formats in Hadoop:

Text Input Format: Default input format in Hadoop.

Key Value Input Format: used for plain text files where the files are broken into lines

Sequence File Input Format: used for reading files in sequence

52)What is Speculative Execution in Hadoop?

Answer)One limitation of Hadoop is that by distributing the tasks on several nodes, there are chances that few slow nodes limit the rest of the program. Tehre are various reasons for the tasks to be slow, which are sometimes not easy to detect. Instead of identifying and fixing the slow-running tasks, Hadoop tries to detect when the task runs slower than expected and then launches other equivalent task as backup. This backup mechanism in Hadoop is Speculative Execution.

It creates a duplicate task on another disk. The same input can be processed multiple times in parallel. When most tasks in a job comes to completion, the speculative execution mechanism schedules duplicate copies of remaining tasks (which are slower) across the nodes that are free currently. When these tasks finish, it is intimated to the JobTracker. If other copies are executing speculatively, Hadoop notifies the TaskTrackers to quit those tasks and reject their output.

53)What is Fault Tolerance?

Answer)Suppose you have a file stored in a system, and due to some technical problem that file gets destroyed. Then there is no chance of getting the data back present in that file. To avoid such situations, Hadoop has introduced the feature of fault tolerance in HDFS. In Hadoop, when we store a file, it automatically gets replicated at two other locations also. So even if one or two of the systems collapse, the file is still available on the third system.

54)What is a heartbeat in HDFS?

Answer)A heartbeat is a signal indicating that it is alive. A datanode sends heartbeat to Namenode and task tracker will send its heart beat to job tracker. If the Namenode or job tracker does not receive heart beat then they will decide that there is some problem in datanode or task tracker is unable to perform the assigned task.

www.smartdatacamp.com

55)How to keep HDFS cluster balanced?

Answer) When copying data into HDFS, it's important to consider cluster balance. HDFS works best when the file blocks are evenly spread across the cluster, so you want to ensure that distcp doesn't disrupt this. For example, if you specified -m 1, a single map would do the copy, which — apart from being slow and not using the cluster resources efficiently — would mean that the first replica of each block would reside on the node running the map (until the disk filled up). The second and third replicas would be spread across the cluster, but this one node would be unbalanced. By having more maps than nodes in the cluster, this problem is avoided. For this reason, it's best to start by running distcp with the default of 20 maps per node.+

However, it's not always possible to prevent a cluster from becoming unbalanced. Perhaps you want to limit the number of maps so that some of the nodes can be used by other jobs. In this case, you can use the balancer tool (see Balancer) to subsequently even out the block distribution across the cluster.

56)How to deal with small files in Hadoop?

Answer)Hadoop Archives (HAR) offers an effective way to deal with the small files problem.

Hadoop Archives or HAR is an archiving facility that packs files in to HDFS blocks efficiently and hence HAR can be used to tackle the small files problem in Hadoop. HAR is created from a collection of files and the archiving tool (a simple command) will run a MapReduce job to process the input files in parallel and create an archive file.

HAR command

hadoop archive -archiveName myhar.har /input/location /output/location

Once a .har file is created, you can do a listing on the .har file and you will see it is made up of index files and part files. Part files are nothing but the original files concatenated together in to a big file. Index files are look up files which is used to look up the individual small files inside the big part files.
hadoop fs -ls /output/location/myhar.har
/output/location/myhar.har/_index
/output/location/myhar.har/_masterindex
/output/location/myhar.har/part-0

57) How to copy file from HDFS to the local file system . There is no physical location of a file under the file , not even directory?

Answer)bin/hadoop fs -get /hdfs/source/path /localfs/destination/path

bin/hadoop fs -copyToLocal /hdfs/source/path /localfs/destination/path

Point your web browser to HDFS WEBUI(namenode_machine:50070), browse to the file you intend to copy, scroll down the page and click on download the file.

58) What's the difference between “hadoop fs” shell commands and “hdfs dfs” shell commands? Are they supposed to be equal? but, why the "hadoop fs" commands show the hdfs files while the "hdfs dfs" commands show the local files?

Answer)Following are the three commands which appears same but have minute differences
hadoop fs {args}
hadoop dfs {args}
hdfs dfs {args}
hadoop fs {args}

FS relates to a generic file system which can point to any file systems like local, HDFS etc. So this can be used when you are dealing with different file systems such as Local FS, HFTP FS, S3 FS, and others

hadoop dfs {args}

dfs is very specific to HDFS. would work for operation relates to HDFS. This has been deprecated and we should use hdfs dfs instead.

hdfs dfs {args}

same as 2nd i.e would work for all the operations related to HDFS and is the recommended command instead of hadoop dfs

below is the list categorized as HDFS commands.

#hdfs commands

namenode|secondarynamenode|datanode|dfs|dfsadmin|fsck|balancer|fetchdt|oiv|dfsgroups

So even if you use Hadoop dfs , it will look locate hdfs and delegate that command to hdfs dfs

59)How to check HDFS Directory size?

Answer)Prior to 0.20.203, and officially deprecated in 2.6.0:
hadoop fs -dus [directory]
Since 0.20.203 (dead link) 1.0.4 and still compatible through 2.6.0:

hdfs dfs -du [-s] [-h] URI [URI ...]
You can also run hadoop fs -help for more info and specifics.

60)Why is there no 'hadoop fs -head' shell command?A fast method for inspecting files on HDFS is to use tail:

~$ hadoop fs -tail /path/to/file

This displays the last kilobyte of data in the file, which is extremely helpful. However, the opposite command head does not appear to be part of the shell command collections. I find this very surprising.

www.smartdatacamp.com

My hypothesis is that since HDFS is built for very fast streaming reads on very large files, there is some access-oriented issue that affects head.

Answer) I would say it's more to do with efficiency - a head can easily be replicated by piping the output of a hadoop fs -cat through the linux head command.

hadoop fs -cat /path/to/file | head

This is efficient as head will close out the underlying stream after the desired number of lines have been output

Using tail in this manner would be considerably less efficient - as you'd have to stream over the entire file (all HDFS blocks) to find the final x number of lines.

hadoop fs -cat /path/to/file | tail

The hadoop fs -tail command as you note works on the last kilobyte - hadoop can efficiently find the last block and skip to the position of the final kilobyte, then stream the output. Piping via tail can't easily do this

61) Difference between hadoop fs -put and hadoop fs -copyFromLocal?

Answer)copyFromLocal is similar to put command, except that the source is restricted to a local file reference.

So, basically you can do with put, all that you do with copyFromLocal, but not vice-versa.

Similarly,

copyToLocal is similar to get command, except that the destination is restricted to a local file reference.

Hence, you can use get instead of copyToLocal, but not the other way round.

62)Is their any HDFS free space available command? Is there a hdfs command to see available free space in hdfs. We can see that through browser at master:hdfsport in browser , but for some reason I can't access this and I need some command. I can see my disk usage through command ./bin/hadoop fs -du -h but cannot see free space available.

Answer) hdfs dfsadmin -report
With older versions of Hadoop, try this:
hadoop dfsadmin -report

63)The default data block size of HDFS/hadoop is 64MB. The block size in disk is generally 4KB. What does 64MB block size mean? ->Does it mean that the smallest unit of read from disk is 64MB?

If yes, what is the advantage of doing that?-> easy for continuous access of large file in HDFS?

Can we do the same by using the original 4KB block size in disk

Answer)What does 64MB block size mean?

The block size is the smallest unit of data that a file system can store. If you store a file that's 1k or 60Mb, it'll take up one block. Once you cross the 64Mb boundry, you need a second block.

If yes, what is the advantage of doing that?

HDFS is meant to handle large files. Lets say you have a 1000Mb file. With a 4k block size, you'd have to make 256,000 requests to get that file (1 request per block). In HDFS, those requests go across a network and come with a lot of overhead. Each request has to be processed by the Name Node to figure out where that block can be found. That's a lot of traffic! If you use 64Mb blocks, the number of requests goes down to 16, greatly reducing the cost of overhead and load on the Name Node.

64)How to specify username when putting files on HDFS from a remote machine? I have a Hadoop cluster setup and working under a common default username "user1". I want to put files into hadoop from a remote machine which is not part of the hadoop cluster. I configured hadoop files on the remote machine in a way that when

hadoop dfs -put file1 ...
is called from the remote machine, it puts the file1 on the Hadoop cluster.

the only problem is that I am logged in as "user2" on the remote machine and that doesn't give me the result I expect. In fact, the above code can only be executed on the remote machine as:

hadoop dfs -put file1 /user/user2/testFolder
However, what I really want is to be able to store the file as:

hadoop dfs -put file1 /user/user1/testFolder
If I try to run the last code, hadoop throws error because of access permissions. Is there anyway that I can specify the username within hadoop dfs command?

I am looking for something like:
hadoop dfs -username user1 file1 /user/user1/testFolder

Answer)By default authentication and authorization is turned off in Hadoop.

www.smartdatacamp.com

The user identity that Hadoop uses for permissions in HDFS is determined by running the whoami command on the client system. Similarly, the group names are derived from the output of running groups.

So, you can create a new whoami command which returns the required username and put it in the PATH appropriately, so that the created whoami is found before the actual whoami which comes with Linux is found. Similarly, you can play with the groups command also.

This is a hack and won't work once the authentication and authorization has been turned on.

If you use the HADOOP_USER_NAME env variable you can tell HDFS which user name to operate with. Note that this only works if your cluster isn't using security features (e.g. Kerberos). For example:

HADOOP_USER_NAME=hdfs hadoop dfs -put ..

65)Where HDFS stores files locally by default?

Answer)You need to look in your hdfs-default.xml configuration file for the dfs.data.dir setting. The default setting is: ${hadoop.tmp.dir}/dfs/data and note that the ${hadoop.tmp.dir} is actually in core-default.xml described here.

The configuration options are described here. The description for this setting is:

Determines where on the local filesystem an DFS data node should store its blocks. If this is a comma-delimited list of directories, then data will be stored in all named directories, typically on different devices. Directories that do not exist are ignored.

66)Hadoop has configuration parameter hadoop.tmp.dir which, as per documentation, is A base for other temporary directories. I presume, this path refers to local file system.

I set this value to /mnt/hadoop-tmp/hadoop-${user.name}. After formatting the namenode and starting all services, I see exactly same path created on HDFS.

Does this mean, hadoop.tmp.dir refers to temporary location on HDFS?

Answer)It's confusing, but hadoop.tmp.dir is used as the base for temporary directories locally, and also in HDFS. The document isn't great, but mapred.system.dir is set by default to "${hadoop.tmp.dir}/mapred/system", and this defines the Path on the HDFS where the Map/Reduce framework stores system files.

If you want these to not be tied together, you can edit your mapred-site.xml such that the definition of mapred.system.dir is something that's not tied to ${hadoop.tmp.dir}

There're three HDFS properties which contain hadoop.tmp.dir in their values

dfs.name.dir: directory where namenode stores its metadata, with default value ${hadoop.tmp.dir}/dfs/name.

dfs.data.dir: directory where HDFS data blocks are stored, with default value ${hadoop.tmp.dir}/dfs/data.

fs.checkpoint.dir: directory where secondary namenode store its checkpoints, default value is ${hadoop.tmp.dir}/dfs/namesecondary.

This is why you saw the /mnt/hadoop-tmp/hadoop-${user.name} in your HDFS after formatting namenode.

67)Is it possible to append to HDFS file from multiple clients in parallel? Basically whole question is in the title. I'm wondering if it's possible to append to file located on HDFS from multiple computers simultaneously? Something like storing stream of events constantly produced by multiple processes. Order is not important.

I recall hearing on one of the Google tech presentations that GFS supports such append functionality but trying some limited testing with HDFS (either with regular file append() or with SequenceFile) doesn't seems to work.

Answer)I don't think that this is possible with HDFS. Even though you don't care about the order of the records, you do care about the order of the bytes in the file. You don't want writer A to write a partial record that then gets corrupted by writer B. This is a hard problem for HDFS to solve on its own, so it doesn't.

Create a file per writer. Pass all the files to any MapReduce worker that needs to read this data. This is much simpler and fits the design of HDFS and Hadoop. If non-MapReduce code needs to read this data as one stream then either stream each file sequentially or write a very quick MapReduce job to consolidate the files.

68)I have 1000+ files available in HDFS with a naming convention of 1_fileName.txt to N_fileName.txt. Size of each file is 1024 MB. I need to merge these files in to one (HDFS)with keeping the order of the file. Say 5_FileName.txt should append only after 4_fileName.txt

What is the best and fastest way to perform this operation.Is there any method to perform this merging without copying the actual data between data nodes? For e-g: Get the block locations of this files and create a new entry (FileName) in the Namenode with these block locations

Answer)There is no efficient way of doing this, you'll need to move all the data to one node, then back to HDFS.

A command line scriptlet to do this could be as follows:

hadoop fs -text *_fileName.txt | hadoop fs -put - targetFilename.txt

This will cat all files that match the glob to standard output, then you'll pipe that stream to the put command and output the stream to an HDFS file named targetFilename.txt

The only problem you have is the filename structure you have gone for - if you have fixed width, zeropadded the number part it would be easier, but in it's current state you'll get an unexpected lexigraphic order (1, 10, 100, 1000, 11, 110, etc) rather than numeric order (1,2,3,4, etc). You could work around this by amending the scriptlet to:

hadoop fs -text [0-9]_fileName.txt [0-9][0-9]_fileName.txt \

[0-9][0-9[0-9]_fileName.txt | hadoop fs -put - targetFilename.txt

69)How to list all files in a directory and its subdirectories in hadoop hdfs?I have a folder in hdfs which has two subfolders each one has about 30 subfolders which,finally,each one contains xml files. I want to list all xml files giving only the main folder's path. Locally I can do this with apache commons-io's FileUtils.listFiles(). I have tried this

FileStatus[] status = fs.listStatus(new Path(args[0]));

but it only lists the two first subfolders and it doesn't go further. Is there any way to do this in hadoop?

Answer)You'll need to use the FileSystem object and perform some logic on the resultant FileStatus objects to manually recurse into the subdirectories.

You can also apply a PathFilter to only return the xml files using the listStatus(Path, PathFilter) method

The hadoop FsShell class has examples of this for the hadoop fs -lsr command, which is a recursive ls -

If you are using hadoop 2.* API there are more elegant solutions:

```
Configuration conf = getConf();
Job job = Job.getInstance(conf);
FileSystem fs = FileSystem.get(conf);
//the second boolean parameter here sets the recursion to true
RemoteIterator (LocatedFileStatus) fileStatusListIterator = fs.listFiles(
new Path("path/to/lib"), true);
while(fileStatusListIterator.hasNext()){
LocatedFileStatus fileStatus = fileStatusListIterator.next();
//do stuff with the file like ...
job.addFileToClassPath(fileStatus.getPath());
}
```

70)Is there a simple command for hadoop that can change the name of a file (in the HDFS) from its old name to a new name?

Answer)Use the following : hadoop fs -mv oldname newname

71)Is there a hdfs command to list files in HDFS directory as per timestamp, ascending or descending? By default, hdfs dfs -ls command gives unsorted list of files.

When I searched for answers what I got was a workaround i.e. hdfs dfs -ls /tmp | sort -k6,7. But is there any better way, inbuilt in hdfs dfs commandline?

Answer)No, there is no other option to sort the files based on datetime.

If you are using hadoop version less than 2.7, you will have to use sort -k6,7 as you are doing:

hdfs dfs -ls /tmp | sort -k6,7
And for hadoop 2.7.x ls command , there are following options available :
Usage: hadoop fs -ls [-d] [-h] [-R] [-t] [-S] [-r] [-u] [args]

Options:
-d: Directories are listed as plain files.
-h: Format file sizes in a human-readable fashion (eg 64.0m instead of 67108864).
-R: Recursively list subdirectories encountered.
-t: Sort output by modification time (most recent first).
-S: Sort output by file size.
-r: Reverse the sort order.
-u: Use access time rather than modification time for display and sorting.
So you can easily sort the files:
hdfs dfs -ls -t -R (-r) /tmp

72)How to unzip .gz files in a new directory in hadoop?I have a bunch of .gz files in a folder in hdfs. I want to unzip all of these .gz files to a new folder in hdfs. How should i do this?

Answer)Using Linux command line

Following command worked.

hadoop fs -cat /tmp/Links.txt.gz | gzip -d | hadoop fs -put - /tmp/unzipped/Links.txt
My gzipped file is Links.txt.gz
The output gets stored in /tmp/unzipped/Links.txt

73)I'm using hdfs -put to load a large 20GB file into hdfs. Currently the process runs @ 4mins. I'm trying to improve the write time of loading data into hdfs. I tried utilizing different block sizes to improve write speed but got the below results:

512M blocksize = 4mins;

www.smartdatacamp.com

256M blocksize = 4mins;
128M blocksize = 4mins;
64M blocksize = 4mins;

Does anyone know what the bottleneck could be and other options we could explore to improve performance of the -put cmd?

Answer)20GB / 4minute comes out to about 85MB/sec. That's pretty reasonable throughput to expect from a single drive with all the overhead of HDFS protocol and network. I'm betting that is your bottleneck. Without changing your ingest process, you're not going to be able to make this magically faster.

The core problem is that 20GB is a decent amount of data and that data getting pushed into HDFS as a single stream. You are limited by disk I/O which is pretty lame given you have a large number of disks in a Hadoop cluster.. You've got a while to go to saturate a 10GigE network (and probably a 1GigE, too).

Changing block size shouldn't change this behavior, as you saw. It's still the same amount of data off disk into HDFS.

I suggest you split the file up into 1GB files and spread them over multiple disks, then push them up with -put in parallel. You might want even want to consider splitting these files over multiple nodes if network becomes a bottleneck. Can you change the way you receive your data to make this faster? Obvious splitting the file and moving it around will take time, too.

www.smartdatacamp.com

Apache MapReduce

Apache MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster.

1) How does Hadoop process records split across block boundaries? Suppose a record line is split across two blocks (b1 and b2). The mapper processing the first block (b1) will notice that the last line doesn't have a EOL separator and fetches the remaining of the line from the next block of data (b2).How does the mapper processing the second block (b2) determine that the first record is incomplete and should process starting from the second record in the block (b2)?

Answer)Map Reduece algorithm does not work on physical blocks of the file. It works on logical input splits. Input split depends on where the record was written. A record may span two Mappers.The way HDFS has been set up, it breaks down very large files into large blocks (for example, measuring 128MB), and stores three copies of these blocks on different nodes in the cluster.HDFS has no awareness of the content of these files. A record may have been started in Block-a but end of that record may be present in Block-b.To solve this problem, Hadoop uses a logical representation of the data stored in file blocks, known as input splits. When a MapReduce job client calculates the input splits, it figures out where the first whole record in a block begins and where the last record in the block ends.

The Map-Reduce framework relies on the InputFormat of the job to:

Validate the input-specification of the job.

Split-up the input file(s) into logical InputSplits, each of which is then assigned to an individual Mapper.

Each InputSplit is then assigned to an individual Mapper for processing. Split could be tuple. InputSplit[] getSplits(JobConf job,int numSplits) is the API to take care of these things.

FileInputFormat, which extends InputFormat implemented getSplits() method.

2) In mapreduce each reduce task write its output to a file named part-r-nnnnn where nnnnn is a partition ID associated with the reduce task. How to merge output files after reduce phase

Answer)We can delegate the entire merging of the reduce output files to hadoop by calling:

hadoop fs -getmerge /output/dir/on/hdfs/ /desired/local/output/file.txt

3) Can you set number of map task in Map reduce?

www.smartdatacamp.com

Answer)The number of map tasks for a given job is driven by the number of input splits. For each input split a map task is spawned. So, over the lifetime of a mapreduce job the number of map tasks is equal to the number of input splits.

4) If your Mapreduce Job launches 20 task for 1 job can you limit to 10 task?

Answer)Yes by setting mapreduce.jobtracker.maxtasks.perjob to 10 in mapred-default.xml file

5) What is the default value set for mapreduce.jobtracker.maxtasks.perjob ?

Answer)Default value is -1 indicates that there is no maximum

6) Have you ever faced Container is running beyond memory limits? For example Container [pid=28921,containerID=container_1389136889968_0001_01_000121] is running beyond virtual memory limits. Current usage: 1.2 GB of 1 GB physical memory used; 2.2 GB of 2.1 GB virtual memory used. Killing container. How to handle this issue?

Answer) For our example cluster, we have the minimum RAM for a Container (yarn.scheduler.minimum-allocation-mb) = 2 GB. We'll thus assign 4 GB for Map task Containers, and 8 GB for Reduce tasks Containers.
In mapred-site.xml:
mapreduce.map.memory.mb: 4096
mapreduce.reduce.memory.mb: 8192

Each Container will run JVMs for the Map and Reduce tasks. The JVM heap size should be set to lower than the Map and Reduce memory defined above, so that they are within the bounds of the Container memory allocated by YARN.
In mapred-site.xml:
mapreduce.map.java.opts: -Xmx3072m
mapreduce.reduce.java.opts: -Xmx6144m

The above settings configure the upper limit of the physical RAM that Map and Reduce tasks will use.

7))What is Shuffling and Sorting in Hadoop MapReduce?

Answer)Shuffling in MapReduce

The process of transferring data from the mappers to reducers is known as shuffling i.e. the process by which the system performs the sort and transfers the map output to the reducer as input. So, MapReduce shuffle phase is necessary for the reducers, otherwise, they would not have any input (or input from every mapper). As shuffling can start even before the map phase has finished so this saves some time and completes the tasks in lesser time.

Sorting in MapReduce

The keys generated by the mapper are automatically sorted by MapReduce Framework, i.e. Before starting of reducer, all intermediate key-value pairs in MapReduce that are generated by mapper get sorted by key and not by value. Values passed to each reducer are not sorted; they can be in any order.

8) What steps do you follow in order to improve the performace of Mapreduce Job?

Answer) There are some general guidelines to improve the performance.

If each task takes less than 30-40 seconds, reduce the number of tasks

If a job has more than 1TB of input, consider increasing the block size of the input dataset to 256M or even 512M so that the number of tasks will be smaller.

Number of reduce tasks per a job should be equal to or a bit less than the number of reduce slots in the cluster.

Some more tips :
Configure the cluster properly with right diagnostic tools
Use compression when you are writing intermediate data to disk
Tune number of Map and Reduce tasks as per above tips
Incorporate Combiner wherever it is appropriate
Use Most appropriate data types for rendering Output (Do not use LongWritable when range of output values are in Integer range. IntWritable is right choice in this case)
Reuse Writables
Have right profiling tools

9)What is the purpose of shuffling and sorting phase in the reducer in Map Reduce Programming?

Answer)First of all shuffling is the process of transfering data from the mappers to the reducers, so I think it is obvious that it is necessary for the reducers, since otherwise, they wouldn't be able to have any input (or input from every mapper). Shuffling can start even before the map phase has finished, to save some time. That's why you can see a reduce status greater than 0% (but less than 33%) when the map status is not yet 100%.

Sorting saves time for the reducer, helping it easily distinguish when a new reduce task should start. It simply starts a new reduce task, when the next key in the sorted input data is different than the previous, to put it simply. Each reduce task takes a list of key-value pairs, but it has to call the reduce() method which takes a key-list(value) input, so it has to group values by key. It's

www.smartdatacamp.com

easy to do so, if input data is pre-sorted (locally) in the map phase and simply merge-sorted in the reduce phase (since the reducers get data from many mappers).

Partitioning, that you mentioned in one of the answers, is a different process. It determines in which reducer a (key, value) pair, output of the map phase, will be sent. The default Partitioner uses a hashing on the keys to distribute them to the reduce tasks, but you can override it and use your own custom Partitioner.

10)How do I submit extra content (jars, static files, etc) for Mapreduce job to use during runtime?

Answer)The distributed cache feature is used to distribute large read-only files that are needed by map/reduce jobs to the cluster. The framework will copy the necessary files from a URL on to the slave node before any tasks for the job are executed on that node. The files are only copied once per job and so should not be modified by the application.

11)How do I get my MapReduce Java Program to read the Cluster's set configuration and not just defaults?

Answer)The configuration property files ({core|mapred|hdfs}-site.xml) that are available in the various conf/ directories of your Hadoop installation needs to be on the CLASSPATH of your Java application for it to get found and applied. Another way of ensuring that no set configuration gets overridden by any Job is to set those properties as final

12)How do I get each of a jobs maps to work on one complete input-file and not allow the framework to split-up the files?

Answer) Essentially a job's input is represented by the InputFormat[interface] or FileInputFormat[base class].

For this purpose one would need a non-splittable FileInputFormat i.e. an input-format which essentially tells the map-reduce framework that it cannot be split-up and processed. To do this you need your particular input-format to return false for the isSplittable call.

E.g.
org.apache.hadoop.mapred.SortValidator.RecordStatsChecker.NonSplitableSequenceFileInputFormat in src/test/org/apache/hadoop/mapred/SortValidator.java

In addition to implementing the InputFormat interface and having isSplitable() returning false, it is also necessary to implement the RecordReader interface for returning the whole content of the input file. Default is LineRecordReader, which splits the file into separate lines

13)You see a maximum of 2 maps/reduces spawned concurrently on each TaskTracker, how do you increase that?

Answer) Use the configuration knob: mapred.tasktracker.map.tasks.maximum and mapred.tasktracker.reduce.tasks.maximum to control the number of maps/reduces spawned simultaneously on a TaskTracker. By default, it is set to 2, hence one sees a maximum of 2 maps and 2 reduces at a given instance on a TaskTracker.

14)How do Map/Reduce InputSplit's handle record boundaries correctly?

Answer)It is the responsibility of the InputSplit's RecordReader to start and end at a record boundary. For SequenceFile's every 2k bytes has a 20 bytes sync mark between the records. These sync marks allow the RecordReader to seek to the start of the InputSplit, which contains a file, offset and length and find the first sync mark after the start of the split. The RecordReader continues processing records until it reaches the first sync mark after the end of the split. The first split of each file naturally starts immediately and not after the first sync mark. In this way, it is guaranteed that each record will be processed by exactly one mapper.

Text files are handled similarly, using newlines instead of sync marks.

15) How do I change final output file name with the desired name rather than in partitions like part-00000, part-00001?

Answer)You can subclass the OutputFormat.java class and write your own. You can locate and browse the code of TextOutputFormat, MultipleOutputFormat.java, etc. for reference. It might be the case that you only need to do minor changes to any of the existing Output Format classes. To do that you can just subclass that class and override the methods you need to change.

16)When writing a New InputFormat, what is the format for the array of string returned by InputSplit\#getLocations()?

Answer)It appears that DatanodeID.getHost() is the standard place to retrieve this name, and the machineName variable, populated in DataNode.java\#startDataNode, is where the name is first set. The first method attempted is to get "slave.host.name" from the configuration; if that is not available, DNS.getDefaultHost is used instead.

17)How do you gracefully stop a running job?

Answer)hadoop job -kill JOBID

18)How do I limit Limiting Task Slot Usage

www.smartdatacamp.com

Answer)There are many reasons why one wants to limit the number of running tasks.

Job is consuming all task slots

The most common reason is because a given job is consuming all of the available task slots, preventing other jobs from running. The easiest and best solution is to switch from the default FIFO scheduler to another scheduler, such as the FairShareScheduler or the CapacityScheduler. Both solve this problem in slightly different ways. Depending upon need, one may be a better fit than the other.

Job has taken too many reduce slots that are still waiting for maps to finish

There is a job tunable called mapred.reduce.slowstart.completed.maps that sets the percentage of maps that must be completed before firing off reduce tasks. By default, this is set to 5% (0.05) which for most shared clusters is likely too low. Recommended values are closer to 80% or higher (0.80). Note that for jobs that have a significant amount of intermediate data, setting this value higher will cause reduce slots to take more time fetching that data before performing work.

Job is referencing an external, limited resource (such as a database)

In Hadoop terms, we call this a 'side-effect'.

One of the general assumptions of the framework is that there are not any side-effects. All tasks are expected to be restartable and a side-effect typically goes against the grain of this rule.

If a task absolutely must break the rules, there are a few things one can do:

Disable SpeculativeExecution .

Deploy ZooKeeper and use it as a persistent lock to keep track of how many tasks are running concurrently

Use a scheduler with a maximum task-per-queue feature and submit the job to that queue, such as FairShareScheduler or CapacityScheduler

19)How to increase the number of slots used?

Answer)There are both job and server-level tunables that impact how many tasks are run concurrently.

Increase the amount of tasks per node

There are two server tunables that determine how many tasks a given TaskTracker will run on a node:
mapred.tasktracker.map.tasks.maximum sets the map slot usage
mapred.tasktracker.reduce.tasks.maximum sets the reduce slot usage

These must be set in the mapred-site.xml file on the TaskTracker. After making the change, the TaskTracker must be restarted to see it. One should see the values increase (or decrease) on the JobTracker main page. Note that this is not set by your job.

Increase the amount of map tasks

Typically, the amount of maps per job is determined by Hadoop based upon the InputFormat and the block size in place. Using mapred.min.split.size and mapred.max.split.size settings, one can provide hints to the system that it should use a size that is different than the block size to determine what the min and max input size should be.

Increase the amount of reduce tasks

Currently, the number of reduces is determined by the job. mapred.reduce.tasks should be set by the job to the appropriate number of reduces. When using Pig, use the PARALLEL keyword.

20)When do reduce tasks start in Hadoop?Do they start after a certain percentage (threshold) of mappers complete? If so, is this threshold fixed? What kind of threshold is typically used?

Answer)The reduce phase has 3 steps: shuffle, sort, reduce. Shuffle is where the data is collected by the reducer from each mapper. This can happen while mappers are generating data since it is only a data transfer. On the other hand, sort and reduce can only start once all the mappers are done. You can tell which one MapReduce is doing by looking at the reducer completion percentage: 0-33% means its doing shuffle, 34-66% is sort, 67%-100% is reduce. This is why your reducers will sometimes seem "stuck" at 33%- it's waiting for mappers to finish.

Reducers start shuffling based on a threshold of percentage of mappers that have finished. You can change the parameter to get reducers to start sooner or later.

Why is starting the reducers early a good thing? Because it spreads out the data transfer from the mappers to the reducers over time, which is a good thing if your network is the bottleneck.

Why is starting the reducers early a bad thing? Because they "hog up" reduce slots while only copying data and waiting for mappers to finish. Another job that starts later that will actually use the reduce slots now can't use them.

You can customize when the reducers startup by changing the default value of mapred.reduce.slowstart.completed.maps in mapred-site.xml. A value of 1.00 will wait for all the mappers to finish before starting the reducers. A value of 0.0 will start the reducers right away. A value of 0.5 will start the reducers when half of the mappers are complete. You can also change mapred.reduce.slowstart.completed.maps on a job-by-job basis. In new versions of Hadoop (at least 2.4.1) the parameter is called is mapreduce.job.reduce.slowstart.completedmaps

Typically, I like to keep mapred.reduce.slowstart.completed.maps above 0.9 if the system ever has multiple jobs running at once. This way the job doesn't hog up reducers when they aren't doing anything but copying data. If you only ever have one job running at a time, doing 0.1 would probably be appropriate.

21) How do you do chaining of multiple Mapreduce job in Hadoop? In many real-life situations where you apply MapReduce, the final algorithms end up being several MapReduce steps.i.e. Map1 , Reduce1 , Map2 , Reduce2 , and so on.So you have the output from the last reduce that is needed as the input for the next map.The intermediate data is something you (in general) do not want to keep once the pipeline has been successfully completed. Also because this intermediate data is in general some data structure (like a

'map' or a 'set') you don't want to put too much effort in writing and reading these key-value pairs.What is the recommended way of doing that in Hadoop?

Answer)You use the JobClient.runJob(). The output path of the data from the first job becomes the input path to your second job. These need to be passed in as arguments to your jobs with appropriate code to parse them and set up the parameters for the job.

I think that the above method might however be the way the now older mapred API did it, but it should still work. There will be a similar method in the new mapreduce API but i'm not sure what it is.

As far as removing intermediate data after a job has finished you can do this in your code. The way i've done it before is using something like:

FileSystem.delete(Path f, boolean recursive);

Where the path is the location on HDFS of the data. You need to make sure that you only delete this data once no other job requires it.

There are many ways you can do it.

(1) Cascading jobs

Create the JobConf object "job1" for the first job and set all the parameters with "input" as inputdirectory and "temp" as output directory. Execute this job:

JobClient.run(job1).

Immediately below it, create the JobConf object "job2" for the second job and set all the parameters with "temp" as inputdirectory and "output" as output directory. Execute this job:

JobClient.run(job2).

(2) Create two JobConf objects and set all the parameters in them just like (1) except that you don't use JobClient.run.

Then create two Job objects with jobconfs as parameters:
Job job1=new Job(jobconf1);
Job job2=new Job(jobconf2);
Using the jobControl object, you specify the job dependencies and then run the jobs:
JobControl jbcntrl=new JobControl("jbcntrl");
jbcntrl.addJob(job1);
jbcntrl.addJob(job2);
job2.addDependingJob(job1);
jbcntrl.run();

(3) If you need a structure somewhat like Map+ | Reduce | Map*, you can use the ChainMapper and ChainReducer classes that come with Hadoop version 0.19 and onwards.

22)Can you explain me how secondary sorting works in hadoop ? Why must one use GroupingComparator and how does it work in hadoop ?

Answer)Grouping Comparator

Once the data reaches a reducer, all data is grouped by key. Since we have a composite key, we need to make sure records are grouped solely by the natural key. This is accomplished by writing a custom GroupPartitioner. We have a Comparator object only considering the yearMonth field of the TemperaturePair class for the purposes of grouping the records together.

```
public class YearMonthGroupingComparator extends WritableComparator {
public YearMonthGroupingComparator() {
super(TemperaturePair.class, true);
}
@Override
public int compare(WritableComparable tp1, WritableComparable tp2) {
TemperaturePair temperaturePair = (TemperaturePair) tp1;
TemperaturePair temperaturePair2 = (TemperaturePair) tp2;
return temperaturePair.getYearMonth().compareTo(temperaturePair2.getYearMonth());
}
}
```

Here are the results of running our secondary sort job:

```
new-host-2:sbin bbejeck$ hdfs dfs -cat secondary-sort/part-r-00000


190101 -206
190102 -333
190103 -272
190104 -61
190105 -33
190106 44
190107 72
190108 44
```

While sorting data by value may not be a common need, it's a nice tool to have in your back pocket when needed. Also, we have been able to take a deeper look at the inner workings of Hadoop by working with custom partitioners and group partitioners.

23)Explain about the basic parameters of mapper and reducer function.

Answer)Mapper Function Parameters

The basic parameters of a mapper function are LongWritable, text, text and IntWritable.

LongWritable, text- Input Parameters

Text, IntWritable- Intermediate Output Parameters

Here is a sample code on the usage of Mapper function with basic parameters –

```
public static class Map extends MapReduceBase implements Mapper {
private final static IntWritable one = new IntWritable (1);
private Text word = new Text () ;}
```

www.smartdatacamp.com

Reducer Function Parameters

The basic parameters of a reducer function are text, IntWritable, text, IntWritable

First two parameters Text, IntWritable represent Intermediate Output Parameters

The next two parameters Text, IntWritable represent Final Output Parameters

24)How data is spilt in Hadoop?

Answer)The InputFormat used in the MapReduce job create the splits. The number of mappers are then decided based on the number of splits. Splits are not always created based on the HDFS block size. It all depends on the programming logic within the getSplits () method of InputFormat.

25)What is the fundamental difference between a MapReduce Split and a HDFS block?

Answer)MapReduce split is a logical piece of data fed to the mapper. It basically does not contain any data but is just a pointer to the data. HDFS block is a physical piece of data.

26) When is it not recommended to use MapReduce paradigm for large scale data processing?

Answer)It is not suggested to use MapReduce for iterative processing use cases, as it is not cost effective, instead Apache Pig can be used for the same.

27) What happens when a DataNode fails during the write process?

Answer)When a DataNode fails during the write process, a new replication pipeline that contains the other DataNodes opens up and the write process resumes from there until the file is closed. NameNode observes that one of the blocks is under-replicated and creates a new replica asynchronously.

28) List the configuration parameters that have to be specified when running a MapReduce job.

Answer)Input and Output location of the MapReduce job in HDFS.

Input and Output Format.

Classes containing the Map and Reduce functions.

JAR file that contains driver classes and mapper, reducer classes.

29) Is it possible to split 100 lines of input as a single split in MapReduce?

Answer)Yes this can be achieved using Class NLineInputFormat

30) Where is Mapper output stored?

Answer)The intermediate key value data of the mapper output will be stored on local file system of the mapper nodes. This directory location is set in the config file by the Hadoop Admin. Once the Hadoop job completes execution, the intermediate will be cleaned up.

31) Explain the differences between a combiner and reducer.

Answer)Combiner can be considered as a mini reducer that performs local reduce task. It runs on the Map output and produces the output to reducers input. It is usually used for network optimization when the map generates greater number of outputs.

Unlike a reducer, the combiner has a constraint that the input or output key and value types must match the output types of the Mapper.

Combiners can operate only on a subset of keys and values i.e. combiners can be executed on functions that are commutative.

Combiner functions get their input from a single mapper whereas reducers can get data from multiple mappers as a result of partitioning.

32) When is it suggested to use a combiner in a MapReduce job?

Answer)Combiners are generally used to enhance the efficiency of a MapReduce program by aggregating the intermediate map output locally on specific mapper outputs. This helps reduce the volume of data that needs to be transferred to reducers. Reducer code can be used as a combiner, only if the operation performed is commutative. However, the execution of a combiner is not assured.

33)What is the relationship between Job and Task in Hadoop?

Answer)A single job can be broken down into one or many tasks in Hadoop.

www.smartdatacamp.com

34)Is it important for Hadoop MapReduce jobs to be written in Java?

Answer)It is not necessary to write Hadoop MapReduce jobs in Java but users can write MapReduce jobs in any desired programming language like Ruby, Perl, Python, R, Awk, etc. through the Hadoop Streaming API.

35) What is the process of changing the split size if there is limited storage space on Commodity Hardware?

Answer)If there is limited storage space on commodity hardware, the split size can be changed by implementing the "Custom Splitter". The call to Custom Splitter can be made from the main method.

36) What are the primary phases of a Reducer?

Answer)The 3 primary phases of a reducer are –
1) Shuffle
2) Sort
3) Reduce

37) What is a TaskInstance?

Answer)The actual Hadoop MapReduce jobs that run on each slave node are referred to as Task instances. Every task instance has its own JVM process. For every new task instance, a JVM process is spawned by default for a task.

38) Can reducers communicate with each other?

Answer)Reducers always run in isolation and they can never communicate with each other as per the Hadoop MapReduce programming paradigm.

39) What is the difference between Hadoop and RDBMS?

Answer)In RDBMS, data needs to be pre-processed being stored, whereas Hadoop requires no pre-processing.

RDBMS is generally used for OLTP processing whereas Hadoop is used for analytical requirements on huge volumes of data.

Database cluster in RDBMS uses the same data files in shared storage whereas in Hadoop the storage is independent of each processing node.

40) Can we search files using wildcards?

Answer)Yes, it is possible to search for file through wildcards.

41) How is reporting controlled in hadoop?

Answer)The file hadoop-metrics.properties file controls reporting.

42) What is the default input type in MapReduce?

Answer) Text

43) Is it possible to rename the output file?

Answer)Yes, this can be done by implementing the multiple format output class.

44) What do you understand by compute and storage nodes?

Answer) Storage node is the system, where the file system resides to store the data for processing.

Compute node is the system where the actual business logic is executed.

45) When should you use a reducer?

Answer) It is possible to process the data without a reducer but when there is a need to combine the output from multiple mappers – reducers are used. Reducers are generally used when shuffle and sort are required.

46) What is the role of a MapReduce partitioner?

www.smartdatacamp.com

Answer)MapReduce is responsible for ensuring that the map output is evenly distributed over the reducers. By identifying the reducer for a particular key, mapper output is redirected accordingly to the respective reducer.

47) What is identity Mapper and identity reducer?

Answer) IdentityMapper is the default Mapper class in Hadoop. This mapper is executed when no mapper class is defined in the MapReduce job.

IdentityReducer is the default Reducer class in Hadoop. This mapper is executed when no reducer class is defined in the MapReduce job. This class merely passes the input key value pairs into the output directory.

48) What do you understand by the term Straggler ?

Answer) A map or reduce task that takes unsually long time to finish is referred to as straggler.

49) What is a MapReduce Combiner?

Answer)Also known as semi-reducer, Combiner is an optional class to combine the map out records using the same key. The main function of a combiner is to accept inputs from Map Class and pass those key-value pairs to Reducer class

50)What is RecordReader in a Map Reduce?

Answer)RecordReader is used to read key/value pairs form the InputSplit by converting the byte-oriented view and presenting record-oriented view to Mapper.

51)What is OutputCommitter?

Answer)OutPutCommitter describes the commit of MapReduce task. FileOutputCommitter is the default available class available for OutputCommitter in MapReduce. It performs the following operations:

Create temporary output directory for the job during initialization.

Then, it cleans the job as in removes temporary output directory post job completion.

Sets up the task temporary output.

Identifies whether a task needs commit. The commit is applied if required.

JobSetup, JobCleanup and TaskCleanup are important tasks during output commit.

52)Explain what happens when Hadoop spawned 50 tasks for a job and one of the task failed?

Answer)It will restart the task again on some other TaskTracker if the task fails more than the defined limit.

53)What is the key difference between Fork/Join and Map/Reduce? Do they differ in the kind of decomposition and distribution (data vs. computation)?

Answer)One key difference is that F-J seems to be designed to work on a single Java VM, while M-R is explicitly designed to work on a large cluster of machines. These are very different scenarios.

F-J offers facilities to partition a task into several subtasks, in a recursive-looking fashion; more tiers, possibility of 'inter-fork' communication at this stage, much more traditional programming. Does not extend (at least in the paper) beyond a single machine. Great for taking advantage of your eight-core.

M-R only does one big split, with the mapped splits not talking between each other at all, and then reduces everything together. A single tier, no inter-split communication until reduce, and massively scalable. Great for taking advantage of your share of the cloud.

54)What type of problems can mapreduce solve?

Answer)For problems requiring processing and generating large data sets. Say running an interest generation query over all accounts a bank hold. Say processing audit data for all transactions that happened in the past year in a bank. The best use case is from Google - generating search index for google search engine.

55)How to get the input file name in the mapper in a Hadoop program?

Answer)First you need to get the input split, using the newer mapreduce API it would be done as follows:

context.getInputSplit();

But in order to get the file path and the file name you will need to first typecast the result into FileSplit.

www.smartdatacamp.com

So, in order to get the input file path you may do the following:

```
Path filePath = ((FileSplit) context.getInputSplit()).getPath();
String filePathString = ((FileSplit) context.getInputSplit()).getPath().toString();
```

Similarly, to get the file name, you may just call upon getName(), like this:

```
String fileName = ((FileSplit) context.getInputSplit()).getPath().getName();
```

56)What is the difference between Hadoop Map Reduce and Google Map Reduce?

Answer)Google MapReduce and Hadoop are two different implementations (instances) of the MapReduce framework/concept. Hadoop is open source , Google MapReduce is not and actually there are not so many available details about it.

Since they work with large data sets, they have to rely on distributed file systems. Hadoop uses as a standard distributed file system the HDFS (Hadoop Distributed File Systems) while Google MapReduce uses GFS (Google File System)

Hadooop is implemented in java. Google MapReduce seems to be in C++.

Apache Hive

Apache Hive data warehouse software facilitates reading, writing, and managing large datasets residing in distributed storage using SQL.

1) What is the definition of Hive? What is the present version of Hive and explain about ACID transactions in Hive?

Answer) Hive is an open source data warehouse system. We can use Hive for analyzing and querying in large data sets of Hadoop files. Its similar to SQL. Hive supports ACID transactions: The full form of ACID is Atomicity, Consistency, Isolation, and Durability. ACID transactions are provided at the row levels, there are Insert, Delete, and Update options so that Hive supports ACID transaction. Insert
Delete
Update

2)Explain what is a Hive variable. What do we use it for?

Answer)Hive variable is basically created in the Hive environment that is referenced by Hive scripting languages. It provides to pass some values to the hive queries when the query starts executing. It uses the source command.

3)What kind of data warehouse application is suitable for Hive? What are the types of tables in Hive?

Answer) Hive is not considered as a full database. The design rules and regulations of Hadoop and HDFS put restrictions on what Hive can do.Hive is most suitable for data warehouse applications.
Where
Analyzing the relatively static data.
Less Responsive time.
No rapid changes in data.Hive doesnot provide fundamental features required for OLTP, Online Transaction Processing.Hive is suitable for data warehouse applications in large data sets. Two types of tables in Hive
Managed table.
External table.

4)How to change the warehouse.dir location for older tables?

www.smartdatacamp.com

Answer)To change the base location of the Hive tables, edit the hive.metastore.warehouse.dir param. This will not affect the older tables. Metadata needs to be changed in the database (MySQL or Derby). The location of Hive tables is in table SDS and column LOCATION.

5)What is Hive Metastore?

Answer)Hive metastore is a database that stores metadata about your Hive tables (eg. tablename, column names and types, table location, storage handler being used, number of buckets in the table, sorting columns if any, partition columns if any, etc.). When you create a table, this metastore gets updated with the information related to the new table which gets queried when you issue queries on that table.

6)Wherever Different Directory I run hive query, it creates new metastore_db, please explain the reason for it?

Answer)Whenever you run the hive in embedded mode, it creates the local metastore. And before creating the metastore it looks whether metastore already exist or not. This property is defined in configuration file hive-site.xml. Property is [javax.jdo.option.ConnectionURL] with default value jdbc:derby:;databaseName=metastore_db;create=true. So to change the behavior change the location to absolute path, so metastore will be used from that location.

7)Is it possible to use same metastore by multiple users, in case of embedded hive?

Answer)No, it is not possible to use metastore in sharing mode. It is recommended to use standalone real database like MySQL or PostGresSQL.

8)Is multiline comment supported in Hive Script ?

Answer) No.

9)If you run hive as a server, what are the available mechanism for connecting it from application?

Answer)There are following ways by which you can connect with the Hive Server

1. Thrift Client: Using thrift you can call hive commands from a various programming languages e.g. C++, Java, PHP, Python and Ruby.

2. JDBC Driver : It supports the Type 4 (pure Java) JDBC Driver

3. ODBC Driver: It supports ODBC protocol.

10)What is SerDe in Apache Hive ?

Answer)A SerDe is a short name for a Serializer Deserializer. Hive uses SerDe and FileFormat to read and write data from tables. An important concept behind Hive is that it DOES NOT own the Hadoop File System format that data is stored in. Users are able to write files to HDFS with whatever tools or mechanism takes their fancy (CREATE EXTERNAL TABLE or LOAD DATA INPATH) and use Hive to correctly parse that file format in a way that can be used by Hive. A SerDe is a powerful and customizable mechanism that Hive uses to parse data stored in HDFS to be used by Hive.

11)Which classes are used by the Hive to Read and Write HDFS Files

Answer) Following classes are used by Hive to read and write HDFS files

TextInputFormat or HiveIgnoreKeyTextOutputFormat: These 2 classes read/write data in plain text file format.

SequenceFileInputFormat or SequenceFileOutputFormat: These 2 classes read/write data in hadoop SequenceFile format.

12)Give examples of the SerDe classes which hive uses to Serialize and Deserilize data ?

Answer)Hive currently use these SerDe classes to serialize and deserialize data:

MetadataTypedColumnsetSerDe: This SerDe is used to read/write delimited records like CSV, tab-separated control-A separated records (quote is not supported yet.)

ThriftSerDe: This SerDe is used to read or write thrift serialized objects. The class file for the Thrift object must be loaded first.

DynamicSerDe: This SerDe also read or write thrift serialized objects, but it understands thrift DDL so the schema of the object can be provided at runtime. Also it supports a lot of different protocols, including TBinaryProtocol, TJSONProtocol, TCTLSeparatedProtocol(which writes data in delimited records).

13)How do you write your own custom SerDe ?

Answer)In most cases, users want to write a Deserializer instead of a SerDe, because users just want to read their own data format instead of writing to it.

For example, the RegexDeserializer will deserialize the data using the configuration parameter regex, and possibly a list of column names

www.smartdatacamp.com

If your SerDe supports DDL (basically, SerDe with parameterized columns and column types), you probably want to implement a Protocol based on DynamicSerDe, instead of writing a SerDe from scratch. The reason is that the framework passes DDL to SerDe through thrift DDL format, and its non-trivial to write a thrift DDL parser.

14)What is ObjectInspector functionality ?

Answer)Hive uses ObjectInspector to analyze the internal structure of the row object and also the structure of the individual columns.

ObjectInspector provides a uniform way to access complex objects that can be stored in multiple formats in the memory, including:

Instance of a Java class (Thrift or native Java)

A standard Java object (we use java.util.List to represent Struct and Array, and use java.util.Map to represent Map)

A lazily-initialized object (For example, a Struct of string fields stored in a single Java string object with starting offset for each field)

A complex object can be represented by a pair of ObjectInspector and Java Object. The ObjectInspector not only tells us the structure of the Object, but also gives us ways to access the internal fields inside the Object.

15)What is the functionality of Query Processor in Apache Hive ?

Answer)This component implements the processing framework for converting SQL to a graph of map or reduce jobs and the execution time framework to run those jobs in the order of dependencies.

16)What is the limitation of Derby database for Hive metastore?

Answer)With derby database, you cannot have multiple connections or multiple sessions instantiated at the same time. Derby database runs in the local mode and it creates a log file so that multiple users cannot access Hive simultaneously.

17)What are managed and external tables?

Answer)We have got two things, one of which is data present in the HDFS and the other is the metadata, present in some database.

There are two categories of Hive tables that is Managed and External Tables.

In the Managed tables, both the data and the metadata are managed by Hive and if you drop the managed table, both data and metadata are deleted.

There are some situations where your data will be controlled by some other application and you want to read that data but you must allow Hive to delete that data. In such case, you can create an external table in Hive. In the external table, metadata is controlled by Hive but the actual data will be controlled by some other application. So, when you delete a table accidentally, only the metadata will be lost and the actual data will reside wherever it is.

18)What are the complex data types in Hive?

Answer)MAP: The Map contains a key-value pair where you can search for a value using the key.

STRUCT:A Struct is a collection of elements of different data types. For example, if you take the address, it can have different data types. For example, pin code will be in Integer format.

ARRAY:An Array will have a collection of homogeneous elements. For example, if you take your skillset, you can have N number of skills

UNIONTYPE:It represents a column which can have a value that can belong to any of the data types of your choice.

19)How does partitioning help in the faster execution of queries?

Answer)With the help of partitioning, a subdirectory will be created with the name of the partitioned column and when you perform a query using the WHERE clause, only the particular sub-directory will be scanned instead of scanning the whole table. This gives you faster execution of queries.

20)How to enable dynamic partitioning in Hive?

Answer)Related to partitioning there are two types of partitioning Static and Dynamic. In the static partitioning, you will specify the partition column while loading the data.

Whereas in dynamic partitioning, you push the data into Hive and then Hive decides which value should go into which partition. To enable dynamic partitioning, you have set the below property

set hive.exec.dynamic.parition.mode = nonstrict;

Example: insert overwrite table emp_details_partitioned partition(location)

select * from emp_details;

21)How does bucketing help in the faster execution of queries?

www.smartdatacamp.com

Answer)If you have to join two large tables, you can go for reduce side join. But if both the tables have the same number of buckets or same multiples of buckets and also sorted on the same column there is a possibility of SMBMJ in which all the joins take place in the map phase itself by matching the corresponding buckets. Buckets are basically files that are created inside the HDFS directory.

There are different properties which you need to set for bucket map joins and they are as follows:
set hive.enforce.sortmergebucketmapjoin = false;
set hive.auto.convert.sortmerge.join = false;
set hive.optimize.bucketmapjoin = ture;
set hive.optimize.bucketmapjoin.sortedmerge = true;

22)How to enable bucketing in Hive?

Answer)By default bucketing is disabled in Hive, you can enforce to enable it by setting the below property

set hive.enforce.bucketing = true;

23)Which method has to be overridden when we use custom UDF in Hive?

Answer)Whenever you write a custom UDF in Hive, you have to extend the UDF class and you have to override the evaluate() function.

24)What are the different file formats in Hive?

Answer)There are different file formats supported by Hive
Text File format
Sequence File format
RC file format
Parquet
Avro
ORC

Every file format has its own characteristics and Hive allows you to choose easily the file format which you wanted to use.

25)How is SerDe different from File format in Hive?

Answer)SerDe stands for Serializer and Deserializer. It determines how to encode and decode the field values or the column values from a record that is how you serialize and deserialize the values of a column

But file format determines how records are stored in key value format or how do you retrieve the records from the table

26)What is RegexSerDe?

Answer)Regex stands for a regular expression. Whenever you want to have a kind of pattern matching, based on the pattern matching, you have to store the fields. RegexSerDe is present in org.apache.hadoop.hive.contrib.serde2.RegexSerDe.

In the SerDeproperties, you have to define your input pattern and output fields. For example, you have to get the column values from line xyz/pq@def if you want to take xyz, pq and def separately.

To extract the pattern, you can use:

input.regex = (.*)/(.*)@(.*)

To specify how to store them, you can use

output.format.string = %1$s%2$s%3$s;

27)How is ORC file format optimised for data storage and analysis?

Answer)ORC stores collections of rows in one file and within the collection the row data will be stored in a columnar format. With columnar format, it is very easy to compress, thus reducing a lot of storage cost.

While querying also, it queries the particular column instead of querying the whole row as the records are stored in columnar format.

ORC has got indexing on every block based on the statistics min, max, sum, count on columns so when you query, it will skip the blocks based on the indexing.

28)How to access HBase tables from Hive?

Answer)Using Hive-HBase storage handler, you can access the HBase tables from Hive and once you are connected, you can query HBase using the SQL queries from Hive. You can also join multiple tables in HBase from Hive and retrieve the result.

29)When running a JOIN query, I see out-of-memory errors.?

www.smartdatacamp.com

Answer)This is usually caused by the order of JOIN tables. Instead of [FROM tableA a JOIN tableB b ON], try [FROM tableB b JOIN tableA a ON] NOTE that if you are using LEFT OUTER JOIN, you might want to change to RIGHT OUTER JOIN. This trick usually solve the problem the rule of thumb is, always put the table with a lot of rows having the same value in the join key on the rightmost side of the JOIN.

30)Did you used Mysql as Metatstore and faced errors like com.mysql.jdbc.exceptions.jdbc4. CommunicationsException: Communications link failure ?

Answer)This is usually caused by MySQL servers closing connections after the connection is idling for some time. Run the following command on the MySQL server will solve the problem [set global wait_status=120]

When using MySQL as a metastore I see the error [com.mysql.jdbc.exceptions.MySQLSyntaxErrorException: Specified key was too long; max key length is 767 bytes].

This is a known limitation of MySQL 5.0 and UTF8 databases. One option is to use another character set, such as latin1, which is known to work.

31)Does Hive support Unicode?

Answer)You can use Unicode string on data or comments, but cannot use for database or table or column name.

You can use UTF-8 encoding for Hive data. However, other encodings are not supported (HIVE 7142 introduce encoding for LazySimpleSerDe, however, the implementation is not complete and not address all cases).

32)Are Hive SQL identifiers (e.g. table names, column names, etc) case sensitive?

Answer)No. Hive is case insensitive.

33)What is the best way to load xml data into hive

Answer)The easiest way is to use the Hive XML SerDe (com.ibm.spss.hive.serde2.xml.XmlSerDe), which will allow you to directly import and work with XML data.

34)When Hive is not suitable?

Answer)It does not provide OLTP transactions support only OLAP transactions.If application required OLAP, switch to NoSQL database.HQL queries have higher latency, due to the mapreduce.

35)Mention what are the different modes of Hive?

Answer)Depending on the size of data nodes in Hadoop, Hive can operate in two modes.

These modes are,Local mode and Map reduce mode

36)Mention what is ObjectInspector functionality in Hive?

Answer)ObjectInspector functionality in Hive is used to analyze the internal structure of the columns, rows, and complex objects. It allows to access the internal fields inside the objects.

37)Mention what is (HS2) HiveServer2?

Answer)It is a server interface that performs following functions.

It allows remote clients to execute queries against Hive

Retrieve the results of mentioned queries

Some advanced features Based on Thrift RPC in its latest version include

Multi-client concurrency

Authentication

38)Mention what Hive query processor does?

Answer)Hive query processor convert graph of MapReduce jobs with the execution time framework. So that the jobs can be executed in the order of dependencies.

39)Mention what are the components of a Hive query processor?

Answer)The components of a Hive query processor include,
Logical Plan Generation
Physical Plan Generation
Execution Engine
Operators
UDFs and UDAFs

www.smartdatacamp.com

Optimizer
Parser
Semantic Analyzer
Type Checking

40)Mention if we can name view same as the name of a Hive table?

Answer)No. The name of a view must be unique compared to all other tables and as views present in the same database.

41)Explain how can you change a column data type in Hive?

Answer)You can change a column data type in Hive by using command,

ALTER TABLE table_name CHANGE column_name column_name new_datatype;

42)Mention what is the difference between order by and sort by in Hive?

Answer)SORT BY will sort the data within each reducer. You can use any number of reducers for SORT BY operation.

ORDER BY will sort all of the data together, which has to pass through one reducer. Thus, ORDER BY in hive uses a single

43)Explain when to use explode in Hive?

Answer)Hadoop developers sometimes take an array as input and convert into a separate table row. To convert complex data types into desired table formats, Hive use explode.

44)Mention how can you stop a partition form being queried?

Answer)You can stop a partition form being queried by using the ENABLE OFFLINE clause with ALTER TABLE statement.

45)What is the need for custom Serde?

Answer)Depending on the nature of data the user has, the inbuilt SerDe may not satisfy the format of the data. SO users need to write their own java code to satisfy their data format requirements.

46)What is the default location where hive stores table data?

Answer)hdfs://namenode_server/user/hive/warehouse

47)Is there a date data type in Hive?

Answer)Yes. The TIMESTAMP data types stores date in java.sql.timestamp format

48)Can we run unix shell commands from hive? Give example?

Answer)Yes, using the ! mark just before the command.For example !pwd at hive prompt will list the current directory.

49)Can hive queries be executed from script files? How?

Answer)Using the source command.
Example
Hive> source /path/to/file/file_with_query.hql

50)What is the importance of .hiverc file?

Answer)It is a file containing list of commands needs to run when the hive CLI starts. For example setting the strict mode to be true etc.

51)What are the default record and field delimiter used for hive text files?

Answer)The default record delimiter is −

And the filed delimiters are − \001,\002,\003

52)What do you mean by schema on read?

Answer)The schema is validated with the data when reading the data and not enforced when writing data.

www.smartdatacamp.com

53)How do you list all databases whose name starts with p?

Answer)SHOW DATABASES LIKE p.*

54)What does the USE command in hive do?

Answer)With the use command you fix the database on which all the subsequent hive queries will run.

55)How can you delete the DBPROPERTY in Hive?

Answer)There is no way you can delete the DBPROPERTY.

56)What is the significance of the line?

Answer)set hive.mapred.mode = strict;

57)How do you check if a particular partition exists?

Answer)This can be done with following query

SHOW PARTITIONS table_name PARTITION(partitioned_column=partition_value)

58)Which java class handles the Input record encoding into files which store the tables in Hive?

Answer)org.apache.hadoop.mapred.TextInputFormat

59)Which java class handles the output record encoding into files which result from Hive queries?

Answer)org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat

60)What is the significance of IF EXISTS clause while dropping a table?

Answer)When we issue the command DROP TABLE IF EXISTS table_name

Hive throws an error if the table being dropped does not exist in the first place.

61)When you point a partition of a hive table to a new directory, what happens to the data?

Answer)The data stays in the old location. It has to be moved manually.

62)Write a query to insert a new column(new_col INT) into a hive table (htab) at a position before an existing column (x_col)

Answer)ALTER TABLE table_name
CHANGE COLUMN new_col INT
BEFORE x_col

63)Does the archiving of Hive tables give any space saving in HDFS?

Answer)No. It only reduces the number of files which becomes easier for namenode to manage.

64)While loading data into a hive table using the LOAD DATA clause, how do you specify it is a hdfs file and not a local file ?

Answer)By Omitting the LOCAL CLAUSE in the LOAD DATA statement.

65)If you omit the OVERWRITE clause while creating a hive table,what happens to file which are new and files which already exist?

Answer)The new incoming files are just added to the target directory and the existing files are simply overwritten. Other files whose name does not match any of the incoming files will continue to exist.

If you add the OVERWRITE clause then all the existing data in the directory will be deleted before new data is written.

66)What does the following query do?

INSERT OVERWRITE TABLE employees

PARTITION (country, state)

www.smartdatacamp.com

SELECT ..., se.cnty, se.st

FROM staged_employees se

Answer)It creates partition on table employees with partition values coming from the columns in the select clause. It is called Dynamic partition insert.

67)What is a Table generating Function on hive?

Answer)A table generating function is a function which takes a single column as argument and expands it to multiple column or rows. Example explode()

68)How can Hive avoid mapreduce?

Answer)If we set the property hive.exec.mode.local.auto to true then hive will avoid mapreduce to fetch query results.

69)What is the difference between LIKE and RLIKE operators in Hive?

Answer)The LIKE operator behaves the same way as the regular SQL operators used in select queries.

Example

street_name like %Chi

But the RLIKE operator uses more advance regular expressions which are available in java

Example

street_name RLIKE .*(Chi|Oho).* which will select any word which has either chi or oho in it.

70)Is it possible to create Cartesian join between 2 tables, using Hive?

Answer)No. As this kind of Join can not be implemented in mapreduce

71)As part of Optimizing the queries in Hive, what should be the order of table size in a join query?

Answer)In a join query the smallest table to be taken in the first position and largest table should be taken in the last position.

72)What is the usefulness of the DISTRIBUTED BY clause in Hive?

Answer)It controls how the map output is reduced among the reducers. It is useful in case of streaming data

73)How will you convert the string 51.2 to a float value in the price column?

Answer)Select cast(price as FLOAT)

74)What will be the result when you do cast(abc as INT)?

Answer)Hive will return NULL

75)Can we LOAD data into a view?

Answer)No. A view can not be the target of a INSERT or LOAD statement.

76)What types of costs are associated in creating index on hive tables?

Answer)Indexes occupies space and there is a processing cost in arranging the values of the column on which index is cerated.

77)Give the command to see the indexes on a table.

Answer)SHOW INDEX ON table_name

This will list all the indexes created on any of the columns in the table table_name.

78)What does /*streamtable(table_name)*/ do?

Answer)It is query hint to stream a table into memory before running the query. It is a query optimization Technique.

79)The following statement failed to execute. What can be the cause?

www.smartdatacamp.com

LOAD DATA LOCAL INPATH ${env:HOME}/country/state/
OVERWRITE INTO TABLE address;

Answer)The local inpath should contain a file and not a directory. The $env:HOME is a valid variable available in the hive environment

80)How do you specify the table creator name when creating a table in Hive?

Answer)The TBLPROPERTIES clause is used to add the creator name while creating a table.
The TBLPROPERTIES is added like
TBLPROPERTIES(creator = Joan)

81)Suppose I have installed Apache Hive on top of my Hadoop cluster using default metastore configuration. Then, what will happen if we have multiple clients trying to access Hive at the same time?

Answer)The default metastore configuration allows only one Hive session to be opened at a time for accessing the metastore. Therefore, if multiple clients try to access the metastore at the same time, they will get an error. One has to use a standalone metastore, i.e. Local or remote metastore configuration in Apache Hive for allowing access to multiple clients concurrently.

Following are the steps to configure MySQL database as the local metastore in Apache Hive:

One should make the following changes in hive-site.xml:
javax.jdo.option.ConnectionURL property should be set to
jdbc:mysql://host/dbname?createDataba
seIfNotExist=true.
javax.jdo.option.ConnectionDriverName property should be set to com.mysql.jdbc.Driver.
One should also set the username and password as:
javax.jdo.option.ConnectionUserName is set to desired username.
javax.jdo.option.ConnectionPassword is set to the desired password.

The JDBC driver JAR file for MySQL must be on the Hive classpath, i.e. The jar file should be copied into the Hive lib directory.

Now, after restarting the Hive shell, it will automatically connect to the MySQL database which is running as a standalone metastore.

82)Is it possible to change the default location of a managed table?

Answer)Yes, it is possible to change the default location of a managed table. It can be achieved by using the clause LOCATION [hdfs_path].

83)When should we use SORT BY instead of ORDER BY?

Answer)We should use SORT BY instead of ORDER BY when we have to sort huge datasets because SORT BY clause sorts the data using multiple reducers whereas ORDER BY sorts all of the data together using a single reducer. Therefore, using ORDER BY against a large number of inputs will take a lot of time to execute.

84)What is dynamic partitioning and when is it used?

Answer)In dynamic partitioning values for partition columns are known in the runtime, i.e. It is known during loading of the data into a Hive table.

One may use dynamic partition in following two cases:

Loading data from an existing non-partitioned table to improve the sampling and therefore, decrease the query latency.

When one does not know all the values of the partitions before hand and therefore, finding these partition values manually from a huge data sets is a tedious task.

85)Suppose, I create a table that contains details of all the transactions done by the customers of year 2016: CREATE TABLE transaction_details (cust_id INT, amount FLOAT, month STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY , ;

Now, after inserting 50,000 tuples in this table, I want to know the total revenue generated for each month. But, Hive is taking too much time in processing this query. How will you solve this problem and list the steps that I will be taking in order to do so?

Answer)We can solve this problem of query latency by partitioning the table according to each month. So, for each month we will be scanning only the partitioned data instead of whole data sets.

As we know, we can not partition an existing non-partitioned table directly. So, we will be taking following steps to solve the very problem:

Create a partitioned table, say partitioned_transaction:

CREATE TABLE partitioned_transaction (cust_id INT, amount FLOAT, country STRING) PARTITIONED BY (month STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY , ;

2. Enable dynamic partitioning in Hive:
SET hive.exec.dynamic.partition = true;
SET hive.exec.dynamic.partition.mode = nonstrict;

3. Transfer the data from the non – partitioned table into the newly created partitioned table:

INSERT OVERWRITE TABLE partitioned_transaction PARTITION (month) SELECT cust_id, amount, country, month FROM transaction_details;

www.smartdatacamp.com

Now, we can perform the query using each partition and therefore, decrease the query time.

86)How can you add a new partition for the month December in the above partitioned table?

Answer)For adding a new partition in the above table partitioned_transaction, we will issue the command give below:

ALTER TABLE partitioned_transaction ADD PARTITION (month=Dec) LOCATION /partitioned_transaction;

87)What is the default maximum dynamic partition that can be created by a mapper/reducer? How can you change it?

Answer)By default the number of maximum partition that can be created by a mapper or reducer is set to 100. One can change it by issuing the following command:

SET hive.exec.max.dynamic.partitions.pernode = value

88)I am inserting data into a table based on partitions dynamically. But, I received an error FAILED ERROR IN SEMANTIC ANALYSIS: Dynamic partition strict mode requires at least one static partition column. How will you remove this error?

Answer)To remove this error one has to execute following commands:
SET hive.exec.dynamic.partition = true;
SET hive.exec.dynamic.partition.mode = nonstrict;

89)Suppose, I have a CSV file sample.csv present in temp directory with the following entries:
id first_name last_name email gender ip_address
1 Hugh Jackman hughjackman@cam.ac.uk Male 136.90.241.52
2 David Lawrence dlawrence1@gmail.com Male 101.177.15.130
3 Andy Hall andyhall2@yahoo.com Female 114.123.153.64
4 Samuel Jackson samjackson231@sun.com Male 89.60.227.31
5 Emily Rose rose.emily4@surveymonkey.com Female 119.92.21.19

How will you consume this CSV file into the Hive warehouse using built SerDe?

Answer)SerDe stands for serializer or deserializer. A SerDe allows us to convert the unstructured bytes into a record that we can process using Hive. SerDes are implemented using Java. Hive comes with several built-in SerDes and many other third-party SerDes are also available.

Hive provides a specific SerDe for working with CSV files. We can use this SerDe for the sample.csv by issuing following commands:
CREATE EXTERNAL TABLE sample
(id int, first_name string,
last_name string, email string,
gender string, ip_address string)
ROW FORMAT SERDE org.apache.hadoop.hive.serde2.OpenCSVSerde
STORED AS TEXTFILE LOCATION temp;

Now, we can perform any query on the table sample:

SELECT first_name FROM sample WHERE gender = male;

90)Suppose, I have a lot of small CSV files present in input directory in HDFS and I want to create a single Hive table corresponding to these files. The data in these files are in the format: {id, name, e-mail, country}. Now, as we know, Hadoop performance degrades when we use lots of small files.

So, how will you solve this problem where we want to create a single Hive table for lots of small files without degrading the performance of the system?

Answer)One can use the SequenceFile format which will group these small files together to form a single sequence file. The steps that will be followed in doing so are as follows:
Create a temporary table:
CREATE TABLE temp_table (id INT, name STRING, e-mail STRING, country STRING)
ROW FORMAT FIELDS DELIMITED TERMINATED BY , STORED AS TEXTFILE;
Load the data into temp_table:
LOAD DATA INPATH input INTO TABLE temp_table;
Create a table that will store data in SequenceFile format:
CREATE TABLE sample_seqfile (id INT, name STRING, e-mail STRING, country STRING)
ROW FORMAT FIELDS DELIMITED TERMINATED BY , STORED AS SEQUENCEFILE;
Transfer the data from the temporary table into the sample_seqfile table:
INSERT OVERWRITE TABLE sample SELECT * FROM temp_table;
Hence, a single SequenceFile is generated which contains the data present in all of the input files and therefore, the problem of having lots of small files is finally eliminated.

91)Can We Change settings within Hive Session? If Yes, How?

Answer)Yes we can change the settings within Hive session, using the SET command. It helps to change Hive job settings for an exact query.

Example: The following commands shows buckets are occupied according to the table definition.

hive> SET hive.enforce.bucketing=true;

We can see the current value of any property by using SET with the property name. SET will list all the properties with their values set by Hive.

www.smartdatacamp.com

hive> SET hive.enforce.bucketing;
hive.enforce.bucketing=true

And this list will not include defaults of Hadoop. So we should use the below like

SET -v

It will list all the properties including the Hadoop defaults in the system.

92)Is it possible to add 100 nodes when we have 100 nodes already in Hive? How?

Answer)Yes, we can add the nodes by following the below steps.
Take a new system create a new username and password.
Install the SSH and with master node setup ssh connections.
Add ssh public_rsa id key to the authorized keys file.
Add the new data node host name, IP address and other details in /etc/hosts slaves file
192.168.1.102 slave3.in slave3.
Start the Data Node on New Node.
Login to the new node like suhadoop or ssh -X hadoop@192.168.1.103.
Start HDFS of a newly added slave node by using the following command
./bin/hadoop-daemon.sh start data node.
Check the output of jps command on a new node

93)Explain the concatenation function in Hive with an example?

Answer)Concatenate function will join the input strings.We can specify the N number of strings separated by a comma.
Example:
CONCAT (It,-,is,-,a,-,eLearning,-,provider);
Output:
It-is-a-eLearning-provider
So, every time we set the limits of the strings by -. If it is common for every strings, then Hive provides another command
CONCAT_WS. In this case,we have to specify the set limits of operator first.
CONCAT_WS (-,It,is,a,eLearning,provider);
Output: It-is-a-eLearning-provider.

94)Explain Trim and Reverse function in Hive with examples?

Answer)Trim function will delete the spaces associated with a string.
Example:
TRIM(BHAVESH);
Output:
BHAVESH
To remove the Leading space

LTRIM(BHAVESH);
To remove the trailing space
RTRIM(BHAVESH);
In Reverse function, characters are reversed in the string.
Example:
REVERSE(BHAVESH);
Output:
HSEVAHB

95)How to change the column data type in Hive? Explain RLIKE in Hive?

Answer)We can change the column data type by using ALTER and CHANGE.

The syntax is :

ALTER TABLE table_name CHANGE column_namecolumn_namenew_datatype;

Example: If we want to change the data type of the salary column from integer to bigint in the employee table.

ALTER TABLE employee CHANGE salary salary BIGINT;RLIKE: Its full form is Right-Like and it is a special function in the Hive. It helps to examine the two substrings. i.e, if the substring of A matches with B then it evaluates to true.

Example:
Bhavesh RLIKE ave True
Bhavesh RLIKE ^B.* True (this is a regular expression)

96)Explain process to access sub directories recursively in Hive queries?

Answer)By using below commands we can access sub directories recursively in Hive
hive> Set mapred.input.dir.recursive=true;
hive> Set hive.mapred.supports.subdirectories=true;

Hive tables can be pointed to the higher level directory and this is suitable for the directory structure which is like /data/country/state/city/

97)How to skip header rows from a table in Hive?

Answer)Header records in log files
System=
Version=
Sub-version=

In the above three lines of headers that we do not want to include in our Hive query. To skip header lines from our tables in the Hive,set a table property that will allow us to skip the header lines.

www.smartdatacamp.com

```
CREATE EXTERNAL TABLE employee (
name STRING,
job STRING,
dob STRING,
id INT,
salary INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY STORED AS TEXTFILE
LOCATION /user/data
TBLPROPERTIES(skip.header.line.count=2);
```

98)What is the maximum size of string data type supported by hive? Mention the Hive support binary formats

Answer)The maximum size of string data type supported by hive is 2 GB.

Hive supports the text file format by default and it supports the binary format Sequence files, ORC files, Avro Data files, Parquet files.

Sequence files: Splittable, compressible and row oriented are the general binary format.

ORC files: Full form of ORC is optimized row columnar format files. It is a Record columnar file and column oriented storage file. It divides the table in row split. In each split stores that value of the first row in the first column and followed sub subsequently.

AVRO datafiles: It is same as a sequence file splittable, compressible and row oriented, but except the support of schema evolution and multilingual binding support.

99)What is the precedence order of HIVE configuration?

Answer)We are using a precedence hierarchy for setting the properties

SET Command in HIVE

The command line –hiveconf option
Hive-site.XML
Hive-default.xml
Hadoop-site.xml
Hadoop-default.xml

100)If you run a select * query in Hive, Why does it not run MapReduce?

Answer)The hive.fetch.task.conversion property of Hive lowers the latency of mapreduce overhead and in effect when executing queries like SELECT, FILTER, LIMIT, etc., it skips mapreduce function

101)How Hive can improve performance with ORC format tables?

Answer)We can store the hive data in highly efficient manner in the Optimized Row Columnar file format. It can simplify many Hive file format limitations. We can improve the performance by using ORC files while reading, writing and processing the data.

```
Set hive.compute.query.using.stats-true;
Set hive.stats.dbclass-fs;
CREATE TABLE orc_table (
idint,
name string)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY \:
LINES TERMINATED BY \n
STORES AS ORC;
```

102)What is available mechanism for connecting from applications, when we run hive as a server?

Answer)Thrift Client: Using thrift you can call hive commands from various programming languages. Example: C++, PHP,Java, Python and Ruby.

JDBC Driver: JDBC Driver supports the Type 4 (pure Java) JDBC Driver

ODBC Driver: ODBC Driver supports the ODBC protocol.

103)Explain about the different types of join in Hive?

Answer)HiveQL has 4 different types of joins –

JOIN- Similar to Outer Join in SQL

FULL OUTER JOIN – Combines the records of both the left and right outer tables that fulfil the join condition.

LEFT OUTER JOIN- All the rows from the left table are returned even if there are no matches in the right table.

RIGHT OUTER JOIN-All the rows from the right table are returned even if there are no matches in the left table.

104)How can you configure remote metastore mode in Hive?

Answer)To configure metastore in Hive, hive-site.xml file has to be configured with the below property –

www.smartdatacamp.com

hive.metastore.uris

thrift: //node1 (or IP Address):9083

IP address and port of the metastore host

105)What happens on executing the below query? After executing the below query, if you modify the column how will the changes be tracked?

Answer)Hive> CREATE INDEX index_bonuspay ON TABLE employee (bonus)

AS org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler;

The query creates an index named index_bonuspay which points to the bonus column in the employee table. Whenever the value of bonus is modified it will be stored using an index value.

106)How to load Data from a .txt file to Table Stored as ORC in Hive?

Answer)LOAD DATA just copies the files to hive datafiles. Hive does not do any transformation while loading data into tables.

So, in this case the input file /home/user/test_details.txt needs to be in ORC format if you are loading it into an ORC table.

A possible workaround is to create a temporary table with STORED AS TEXT, then LOAD DATA into it, and then copy data from this table to the ORC table.
Here is an example:
CREATE TABLE test_details_txt(visit_id INT, store_id SMALLINT) STORED AS TEXTFILE;
CREATE TABLE test_details_orc(visit_id INT, store_id SMALLINT) STORED AS ORC;
Load into Text table
LOAD DATA LOCAL INPATH /home/user/test_details.txt INTO TABLE test_details_txt;
Copy to ORC table
INSERT INTO TABLE test_details_orc SELECT * FROM test_details_txt;

107)How to create HIVE Table with multi character delimiter

Answer)FILELDS TERMINATED BY does not support multi-character delimiters. The easiest way to do this is to use RegexSerDe:

```
CREATE EXTERNAL TABLE tableex(id INT, name STRING)
ROW FORMAT org.apache.hadoop.hive.contrib.serde2.RegexSerDe
WITH SERDEPROPERTIES (
input.regex = ^(\\d+)~\\*(.*)$
)
STORED AS TEXTFILE
LOCATION /user/myusername;
```

108)Is there any way to get the column name along with the output while execute any query in Hive?

Answer)If we want to see the columns names of the table in HiveQl, the following hive conf property should be set to true.

hive> set hive.cli.print.header=true;

If you prefer to see the column names always then update the $HOME/.hiverc file with the above setting in the first line..

Hive automatically looks for a file named .hiverc in your HOME directory and runs the commands it contains, if any

109)How to Improve Hive Query Performance With Hadoop?

Answer)Use Tez Engine

Apache Tez Engine is an extensible framework for building high-performance batch processing and interactive data processing. It is coordinated by YARN in Hadoop. Tez improved the MapReduce paradigm by increasing the processing speed and maintaining the MapReduce ability to scale to petabytes of data.

Tez engine can be enabled in your environment by setting hive.execution.engine to tez:

set hive.execution.engine=tez;

Use Vectorization

Vectorization improves the performance by fetching 1,024 rows in a single operation instead of fetching single row each time. It improves the performance for operations like filter, join, aggregation, etc.

Vectorization can be enabled in the environment by executing below commands.
set hive.vectorized.execution.enabled=true;
set hive.vectorized.execution.reduce.enabled=true;

Use ORCFile

Optimized Row Columnar format provides highly efficient ways of storing the hive data by reducing the data storage format by 75% of the original. The ORCFile format is better than the Hive files format when it comes to reading, writing, and processing the data. It uses techniques like predicate push-down, compression, and more to improve the performance of the query.

Use Partitioning

With partitioning, data is stored in separate individual folders on HDFS. Instead of querying the whole dataset, it will query partitioned dataset.
1)Create Temporary Table and Load Data Into Temporary Table
2)Create Partitioned Table
3)Enable Dynamic Hive Partition
4)Import Data From Temporary Table To Partitioned Table

www.smartdatacamp.com

Use Bucketing

The Hive table is divided into a number of partitions and is called Hive Partition. Hive Partition is further subdivided into clusters or buckets and is called bucketing or clustering.

Cost-Based Query Optimization

Hive optimizes each querys logical and physical execution plan before submitting for final execution. However, this is not based on the cost of the query during the initial version of Hive.

During later versions of Hive, query has been optimized according to the cost of the query (like which types of join to be performed, how to order joins, the degree of parallelism, etc.).

www.smartdatacamp.com

Apache Pig

Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.

1)What is Pig?

Answer)Apache Pig is a platform, used to analyze large data sets representing them as data flows. It is designed to provide an abstraction over MapReduce, reducing the complexities of writing a MapReduce task using Java programming. We can perform data manipulation operations very easily in Hadoop using Apache Pig. Apache Pig has two main components – the Pig Latin language and the Pig Run-time Environment, in which Pig Latin programs are executed.

2)How can I pass a specific hadoop configuration parameter to Pig?

Answer)There are multiple places you can pass hadoop configuration parameter to Pig. Here is a list from high priority to low priority (configuration in high priority will override the configuration in low priority):
1. set command
2. -P properties_file
3. pig.properties
4. java system property/environmental variable
5. Hadoop configuration file: hadoop-site.xml/core-site.xml/hdfs-site.xml/mapred-site.xml, or Pig specific hadoop configuration file: pig-cluster-hadoop-site.xml

3)I already register my LoadFunc/StoreFunc jars in "register" statement, but why I still get "Class Not Found" exception?

Answer)Try to put your jars in PIG_CLASSPATH as well. "register" guarantees your jar will be shipped to backend. But in the frontend, you still need to put the jars in CLASSPATH by setting "PIG_CLASSPATH" environment variable.

4)How can I load data using Unicode control characters as delimiters?

Answer)The first parameter to PigStorage is the dataset name, the second is a regular expression to describe the delimiter. We used `String.split(regex, -1)` to extract fields from lines. See java.util.regex.Pattern for more information on the way to use special characters in regex.

www.smartdatacamp.com

If you are loading a file which contains Ctrl+A as separators, you can specify this to PigStorage using the Unicode notation.

LOAD 'input.dat' USING PigStorage('\u0001')as (x,y,z);

5)How do I control the number of mappers?

Answer)It is determined by your InputFormat. If you are using PigStorage, FileInputFormat will allocate at least 1 mapper for each file. If the file is large, FileInputFormat will split the file into smaller trunks. You can control this process by two hadoop setting: "mapred.min.split.size", "mapred.max.split.size". In addition, after InputFormat tells Pig all the splits information, Pig will try to combine small input splits into one mapper. This process can be controlled by "pig.noSplitCombination" and "pig.maxCombinedSplitSize".

6)How do I make my Pig jobs run on a specified number of reducers?

Answer)You can achieve this with the PARALLEL clause.

For example: C = JOIN A by url, B by url PARALLEL 50.

Besides PARALLEL clause, you can also use "set default_parallel" statement in Pig script, or set "mapred.reduce.tasks" system property to specify default parallel to use. If none of these values are set, Pig will only use 1 reducers. (In Pig 0.8, we change the default reducer from 1 to a number calculated by a simple heuristic for foolproof purpose)

7)Can I do a numerical comparison while filtering?

Answer)Yes, you can choose between numerical and string comparison. For numerical comparison use the operators =, and for string comparisons use eq, neq etc.

8)Does Pig support regular expressions?

Answer)Pig does support regular expression matching via the `matches` keyword. It uses java.util.regex matches which means your pattern has to match the entire string (e.g. if your string is `"hi fred"` and you want to find `"fred"` you have to give a pattern of `".*fred"` not `"fred"`).

9)How do I prevent failure if some records don't have the needed number of columns?

Answer)You can filter away those records by including the following in your Pig program:

```
A = LOAD 'foo' USING PigStorage('\t');
B = FILTER A BY ARITY(*) < 5;
```

This code would drop all records that have fewer than five (5) columns.

10)Is there any difference between `==` and `eq` for numeric comparisons?

Answer)There is no difference when using integers. However, `11.0` and `11` will be equal with `==` but not with `eq`.

11)Is there an easy way for me to figure out how many rows exist in a dataset from it's alias?

Answer)You can run the following set of commands, which are equivalent to `SELECT COUNT(*)` in SQL:

```
a = LOAD 'mytestfile.txt';
b = GROUP a ALL;
c = FOREACH b GENERATE COUNT(a.$0);
```

12)Does Pig allow grouping on expressions?

Answer)Pig allows grouping of expressions. For example:

```
grunt> a = LOAD 'mytestfile.txt' AS (x,y,z);
grunt> DUMP a;
(1,2,3)
(4,2,1)
(4,3,4)
(4,3,4)
(7,2,5)
(8,4,3)
b = GROUP a BY (x+y);
(3.0,{(1,2,3)})
(6.0,{(4,2,1)})
(7.0,{(4,3,4),(4,3,4)})
(9.0,{(7,2,5)})
(12.0,{(8,4,3)})
```

If the grouping is based on constants, the result is the same as GROUP ALL except the group-id is replaced by the constant.

```
grunt> b = GROUP a BY 4;
(4,{(1,2,3),(4,2,1),(4,3,4),(4,3,4),(7,2,5),(8,4,3)})
```

13)Is there a way to check if a map is empty?

www.smartdatacamp.com

Answer)In Pig 2.0 you can test the existence of values in a map using the null construct:

m#'key' is not null

14)I load data from a directory which contains different file. How do I find out where the data comes from?

Answer)You can write a LoadFunc which append filename into the tuple you load.

Eg,
A = load '*.txt' using PigStorageWithInputPath();
Here is the LoadFunc:
public class PigStorageWithInputPath extends PigStorage {
Path path = null;

@Override
public void prepareToRead(RecordReader reader, PigSplit split) {
super.prepareToRead(reader, split);
path = ((FileSplit)split.getWrappedSplit()).getPath();
}

@Override
public Tuple getNext() throws IOException {
Tuple myTuple = super.getNext();
if (myTuple != null)
myTuple.append(path.toString());
return myTuple;
}
}

15)How can I calculate a percentage (partial aggregate / total aggregate)?

Answer)The challenge here is to get the total aggregate into the same statement as the partial aggregate. The key is to cast the relation for the total aggregate to a scalar:
A = LOAD 'sample.txt' AS (x:int, y:int);
B = foreach (group A all) generate COUNT(A) as total;
C = foreach (group A by x) generate group as x, (double)COUNT(A) / (double) B.total as percentage;

16)How can I pass a parameter with space to a pig script?

Answer)# Following should work

-p \"NAME='Firstname Lastname'\"

-p \"NAME=Firstname\ Lastname\"

17)What is the difference between logical and physical plans?

Answer)Pig undergoes some steps when a Pig Latin Script is converted into MapReduce jobs by the compiler. Logical and Physical plans are created during the execution of a pig script.

After performing the basic parsing and semantic checking, the parser produces a logical plan and no data processing takes place during the creation of a logical plan. The logical plan describes the logical operators that have to be executed by Pig during execution. For each line in the Pig script, syntax check is performed for operators and a logical plan is created. If an error is encountered, an exception is thrown and the program execution ends.

A logical plan contains a collection of operators in the script, but does not contain the edges between the operators.

After the logical plan is generated, the script execution moves to the physical plan where there is a description about the physical operators, Apache Pig will use, to execute the Pig script. A physical plan is like a series of MapReduce jobs, but the physical plan does not have any reference on how it will be executed in MapReduce.

18)How Pig programming gets converted into MapReduce jobs?

Answer)Pig is a high-level platform that makes many Hadoop data analysis issues easier to execute. A program written in Pig Latin is a data flow language, which need an execution engine to execute the query. So, when a program is written in Pig Latin, Pig compiler converts the program into MapReduce jobs.

19)What are the components of Pig Execution Environment?

Answer)The components of Apache Pig Execution Environment are:

Pig Scripts: Pig scripts are submitted to the Apache Pig execution environment which can be written in Pig Latin using built-in operators and UDFs can be embedded in it.

Parser: The Parser does the type checking and checks the syntax of the script. The parser outputs a DAG (directed acyclic graph). DAG represents the Pig Latin statements and logical operators.

Optimizer: The Optimizer performs the optimization activities like split, merge, transform, reorder operators, etc. The optimizer provides the automatic optimization feature to Apache Pig. The optimizer basically aims to reduce the amount of data in the pipeline.

Compiler: The Apache Pig compiler converts the optimized code into MapReduce jobs automatically.

www.smartdatacamp.com

Execution Engine: Finally, the MapReduce jobs are submitted to the execution engine. Then, the MapReduce jobs are executed and the required result is produced.

20)What are the different ways of executing Pig script?

Answer)There are three ways to execute the Pig script:

Grunt Shell: This is Pig's interactive shell provided to execute all Pig Scripts.

Script File: Write all the Pig commands in a script file and execute the Pig script file. This is executed by the Pig Server.

Embedded Script: If some functions are unavailable in built-in operators, we can programmatically create User Defined Functions (UDF) to bring that functionality using other languages like Java, Python, Ruby, etc. and embed it in the Pig Latin Script file. Then, execute that script file.

21)What are the data types of Pig Latin?

Answer)Pig Latin can handle both atomic data types like int, float, long, double etc. and complex data types like tuple, bag and map.

Atomic or scalar data types are the basic data types which are used in all the languages like string, int, float, long, double, char[], byte[]. These are also called the primitive data types.

The complex data types supported by Pig Latin are:
Tuple: Tuple is an ordered set of fields which may contain different data types for each field.
Bag: A bag is a collection of a set of tuples and these tuples are a subset of rows or entire rows of a table.
Map: A map is key-value pairs used to represent data elements. The key must be a chararray [] and should be unique like column name, so it can be indexed and value associated with it can be accessed on the basis of the keys. The value can be of any data type.

22)Is it possible to pivot a table in one pass in Apache Pig.

Input:
Id Column1 Column2 Column3
1 Row11 Row12 Row13
2 Row21 Row22 Row23
Output:
Id Name Value
1 Column1 Row11
1 Column2 Row12
1 Column3 Row13
2 Column1 Row21
2 Column2 Row22

2 Column3 Row23

Answer)You can do it in 2 ways: 1. Write a UDF which returns a bag of tuples. It will be the most flexible solution, but requires Java code; 2. Write a rigid script like this:
inpt = load '/pig_fun/input/pivot.txt' as (Id, Column1, Column2, Column3);
bagged = foreach inpt generate Id, TOBAG(TOTUPLE('Column1', Column1), TOTUPLE('Column2', Column2), TOTUPLE('Column3', Column3)) as toPivot;
pivoted_1 = foreach bagged generate Id, FLATTEN(toPivot) as t_value;
pivoted = foreach pivoted_1 generate Id, FLATTEN(t_value);
dump pivoted;
Running this script got me following results:
(1,Column1,11)
(1,Column2,12)
(1,Column3,13)
(2,Column1,21)
(2,Column2,22)
(2,Column3,23)
(3,Column1,31)
(3,Column2,32)
(3,Column3,33)

23)How to Load multiple files from a date range (part of the directory structure)I have the following scenario-

Sample HDFS directory structure:
/user/training/test/20100810/data files
/user/training/test/20100811/data files
/user/training/test/20100812/data files
/user/training/test/20100813/data files
/user/training/test/20100814/data files

As you can see in the paths listed above, one of the directory names is a date stamp.

Problem: I want to load files from a date range say from 20100810 to 20100813.

Answer)The path expansion is done by the shell. One common way to solve your problem is to simply use Pig parameters (which is a good way to make your script more resuable anyway):

shell:

pig -f script.pig -param input=/user/training/test/{20100810..20100812}

script.pig:

temp = LOAD '$input' USING SomeLoader() AS ();

24)How to reference columns in a FOREACH after a JOIN?
A = load 'a.txt' as (id, a1);

www.smartdatacamp.com

B = load 'b.txt as (id, b1);
C = join A by id, B by id;
D = foreach C generate id,a1,b1;
dump D;

4th line fails on: Invalid field projection. Projected field [id] does not exist in schema, How to fix this?

Answer)Solution:
A = load 'a.txt' as (id, a1);
B = load 'b.txt as (id, b1);
C = join A by id, B by id;
D = foreach C generate A::id,a1,b1;
dump D;

25)how to include external jar file using PIG

Answer)register /local/path/to/myJar.jar

26)Removing duplicates using PigLatin
Input:
User1 8 NYC
User1 9 NYC
User1 7 LA
User2 4 NYC
User2 3 DC
Output:
User1 8 NYC
User2 4 NYC

Answer)In order to select one record per user (any record) you could use a GROUP BY and a nested FOREACH with LIMIT.
Ex:
inpt = load '.....' .....;
user_grp = GROUP inpt BY $0;
filtered = FOREACH user_grp {
top_rec = LIMIT inpt 1;
GENERATE FLATTEN(top_rec);
};

27)Currently, when we STORE into HDFS, it creates many part files.Is there any way to store out to a single CSV file in Apache Pig?

Answer)You can do this in a few ways:

To set the number of reducers for all Pig opeations, you can use the default_parallel property - but this means every single step will use a single reducer, decreasing throughput:

set default_parallel 1;

Prior to calling STORE, if one of the operations execute is (COGROUP, CROSS, DISTINCT, GROUP, JOIN (inner), JOIN (outer), and ORDER BY), then you can use the PARALLEL 1 keyword to denote the use of a single reducer to complete that command:

GROUP a BY grp PARALLEL 1;

28)I have data that's already grouped and aggregated, it looks like so:
user value count
---- -------- ------
Alice third 5
Alice first 11
Alice second 10
Alice fourth 2
Bob second 20
Bob third 18
Bob first 21
Bob fourth 8

For every user (Alice and Bob), I want retrieve their top n values (let's say 2), sorted terms of 'count'. So the desired output I want is this:

Alice first 11
Alice second 10
Bob first 21
Bob second 20

How can I accomplish this in Apache Pig?

Answer)One approach is
records = LOAD '/user/nubes/ncdc/micro-tab/top.txt' AS
(user:chararray,value:chararray,counter:int);
grpd = GROUP records BY user;
top3 = foreach grpd {
sorted = order records by counter desc;
top = limit sorted 2;
generate group, flatten(top);
};

Input is:
Alice third 5

Alice first 11
Alice second 10
Alice fourth 2
Bob second 20
Bob third 18
Bob first 21
Bob fourth 8
Output is:
(Alice,Alice,first,11)
(Alice,Alice,second,10)
(Bob,Bob,first,21)
(Bob,Bob,second,20)

29)Find if a string is present inside another string in Pig

Answer)You can use this :

```
X = FILTER A BY (f1 matches '.*the_word_you're_looking_for.*');
```

30)how to do Transpose in corresponding few columns in pig?
Input:
id jan feb march
1 j1 f1 m1
2 j2 f2 m2
3 j3 f3 m3
Output:
id value month
1 j1 jan
1 f1 feb
1 m1 march
2 j2 jan
2 f2 feb
2 m2 march
3 j3 jan
3 f3 feb
3 m3 march

Answer)PigScript:

```
A = LOAD 'input.txt' USING PigStorage() AS (id,month1,month2,month3);
B = FOREACH A GENERATE
FLATTEN(TOBAG(TOTUPLE(id,month1,'jan'),TOTUPLE(id,month2,'feb'),TOTUPLE(id,month3,'mar')));
DUMP B;
```

Output:
(1,j1,jan)
(1,f1,feb)

(1,m1,mar)
(2,j2,jan)
(2,f2,feb)
(2,m2,mar)
(3,j3,jan)
(3,f3,feb)
(3,m3,mar)

31)What is the difference between Store and dump commands?

Answer)Dump command after process the data displayed on the terminal, but it's not stored anywhere. Where as Store stored in local file system or HDFS and output execute in a folder. In the protection environment most often hadoop developer used 'store' command to store data in in the HDFS.

32)How to debug a pig script?

Answer)There are several method to debug a pig script. Simple method is step by step execution of a relation and then verify the result. These commands are useful to debug a pig script.

DUMP - Use the DUMP operator to run (execute) Pig Latin statements and display the results to your screen.

ILLUSTRATE - Use the ILLUSTRATE operator to review how data is transformed through a sequence of Pig Latin statements. ILLUSTRATE allows you to test your programs on small datasets and get faster turnaround times.

EXPLAIN - Use the EXPLAIN operator to review the logical, physical, and map reduce execution plans that are used to compute the specified relationship.

DESCRIBE - Use the DESCRIBE operator to view the schema of a relation. You can view outer relations as well as relations defined in a nested FOREACH statement.

33)What are the limitations of the Pig?

Answer)Limitations of the Apache Pig are:

As the Pig platform is designed for ETL-type use cases, it's not a better choice for real-time scenarios.

Apache Pig is not a good choice for pinpointing a single record in huge data sets.

Apache Pig is built on top of MapReduce, which is batch processing oriented.

34)What is BloomMapFile used for?

www.smartdatacamp.com

Answer)The BloomMapFile is a class that extends MapFile. So its functionality is similar to MapFile. BloomMapFile uses dynamic Bloom filters to provide quick membership test for the keys. It is used in Hbase table format.

35)What is the difference between GROUP and COGROUP operators in Pig?

Answer)Group and Cogroup operators are identical. For readability, GROUP is used in statements involving one relation and COGROUP is used in statements involving two or more relations. Group operator collects all records with the same key. Cogroup is a combination of group and join, it is a generalization of a group instead of collecting records of one input depends on a key, it collects records of n inputs based on a key. At a time, we can Cogroup up to 127 relations.

36)Give some list of relational operators used in Pig?

Answer)COGROUP: Joins two or more tables and then perform GROUP operation on the joined table result.
CROSS: CROSS operator is used to compute the cross product (Cartesian product) of two or more relations.
DISTINCT: Removes duplicate tuples in a relation.
FILTER: Select a set of tuples from a relation based on a condition.
FOREACH: Iterate the tuples of a relation, generating a data transformation.
GROUP: Group the data in one or more relations.
JOIN: Join two or more relations (inner or outer join).
LIMIT: Limit the number of output tuples.
LOAD: Load data from the file system.
ORDER: Sort a relation based on one or more fields.
SPLIT: Partition a relation into two or more relations.
STORE: Store data in the file system.
UNION: Merge the content of two relations. To perform a UNION operation on two relations, their columns and domains must be identical.

37)Can we process vast amount of data in local mode? Why?

Answer)No, System has limited fixed amount of storage, where as Hadoop can handle vast amount of data. So Pig -x Mapreduce mode is the best choice to process vast amount of data.

38)Explain about the different complex data types in Pig?

Answer)Apache Pig supports 3 complex data types-

Maps- These are key, value stores joined together using #.

Tuples- Just similar to the row in a table, where different items are separated by a comma. Tuples can have multiple attributes.

Bags- Unordered collection of tuples. Bag allows multiple duplicate tuples.

39)Differentiate between the logical and physical plan of an Apache Pig script?

Answer)Logical and Physical plans are created during the execution of a pig script. Pig scripts are based on interpreter checking. Logical plan is produced after semantic checking and basic parsing and no data processing takes place during the creation of a logical plan. For each line in the Pig script, syntax check is performed for operators and a logical plan is created. Whenever an error is encountered within the script, an exception is thrown and the program execution ends, else for each statement in the script has its own logical plan.

A logical plan contains collection of operators in the script but does not contain the edges between the operators.

After the logical plan is generated, the script execution moves to the physical plan where there is a description about the physical operators, Apache Pig will use, to execute the Pig script. A physical plan is more or less like a series of MapReduce jobs but then the plan does not have any reference on how it will be executed in MapReduce. During the creation of physical plan, cogroup logical operator is converted into 3 physical operators namely –Local Rearrange, Global Rearrange and Package. Load and store functions usually get resolved in the physical plan.

40)What do you understand by an inner bag and outer bag in Pig?

Answer)A relation inside a bag is referred to as inner bag and outer bag is just a relation in Pig

41)Explain the difference between COUNT_STAR and COUNT functions in Apache Pig?

Answer)COUNT function does not include the NULL value when counting the number of elements in a bag, whereas COUNT_STAR (0 function includes NULL values while counting.

42)Explain about the scalar datatypes in Apache Pig.

Answer)integer, float, double, long, bytearray and char array are the available scalar datatypes in Apache Pig.

43)Is it posible to join multiple fields in pig scripts?

www.smartdatacamp.com

Answer) Yes,Join select records from one input and join with another input.This is done by indicating keys for each input. When those keys are equal, the two rows are joined.
input2 = load 'daily' as (exchanges, stocks);
input3 = load 'week' as (exchanges, stocks);
grpds = join input2 by stocks,input3 by stocks;

we can also join multiple keys
example:
input2 = load 'daily' as (exchanges, stocks);
input3 = load 'week' as (exchanges, stocks);
grpds = join input2 by (exchanges,stocks),input3 by (exchanges,stocks);

44)What are the different String functions available in pig?

Answer: Below are most commonly used STRING pig functions
UPPER
LOWER
TRIM
SUBSTRING
INDEXOF
STRSPLIT
LAST_INDEX_OF

45)While writing evaluate UDF, which method has to be overridden?

Answer)While writing UDF in pig, you have to override the method exec() and the base class can be different, while writing filter UDF, you will have to extend FilterFunc and for evaluate UDF, you will have to extend the EvalFunc.EvaluFunc is parameterized and must provide the return type also.

46)What is a skewed join?

Answer)Whenever you want to perform a join with a skewed dataset i.e., a particular value will be repeated many times.

Suppose, if you have two datasets which contains the details about city and the person living in that city. The second dataset contains the details of city and the country.

So automatically city name will be repeated multiple times based on the population of the city and if you want to perform join using the city column then a particular reducer will receive a lot of values for that particular city.

In the skewed dataset, the left input on the join predicate will be divided and even if you have skeweness in the data your data will be split across different machines and the input on the right

side will be duplicated and split across different machines and in this way skewed join is handled in the Pig.

47)Write a word count program in pig?

Answer)lines = LOAD '/user/hadoop/HDFS_File.txt' AS (line:chararray);
words = FOREACH lines GENERATE FLATTEN(TOKENIZE (line)) as word;
grouped = GROUP words by word;
wordcount = FOREACH grouped GENERATE group, COUNT (words);
DUMP wordcount;

48)What is the difference between Pig Latin and HiveQL ?

Answer)Pig Latin:
Pig Latin is a Procedural language
Nested relational data model
Schema is optional
HiveQL:
HiveQL is Declarative
HiveQL flat relational
Schema is required

49)Does Pig support multi-line commands?

Answer)Yes, pig supports both single line and multi-line commands. In single line command it executes the data, but it doesn't store in the file system, but in multiple lines commands it stores the data into '/output';/* , so it can store the data in HDFS.

50)What is the function of UNION and SPLIT operators? Give examples.

Answer)Union operator helps to merge the contents of two or more relations.
Syntax: grunt> Relation_name3 = UNION Relation_name1, Relation_name2
Example: grunt> INTELLIPAAT = UNION intellipaat_data1.txt intellipaat_data2.txt
SPLIT operator helps to divide the contents of two or more relations.
Syntax: grunt> SPLIT Relationa1_name INTO Relationa2_name IF (condition1), Relation2_name (condition2);
Example: SPLIT student_details into student_details1 if marks <35, student_details2 if (8590);

www.smartdatacamp.com

Apache Spark

Apache Spark is a fast and general-purpose cluster computing system. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs. It also supports a rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming.

1)How does Spark relate to Apache Hadoop?

Answer)Spark is a fast and general processing engine compatible with Hadoop data. It can run in Hadoop clusters through YARN or Spark's standalone mode, and it can process data in HDFS, HBase, Cassandra, Hive, and any Hadoop InputFormat. It is designed to perform both batch processing (similar to MapReduce) and new workloads like streaming, interactive queries, and machine learning.

2)Who is using Spark in production?

Answer)As of 2016, surveys show that more than 1000 organizations are using Spark in production. Some of them are listed on the Powered By page and at the Spark Summit.

3)How large a cluster can Spark scale to?

Answer)Many organizations run Spark on clusters of thousands of nodes. The largest cluster we know has 8000 of them. In terms of data size, Spark has been shown to work well up to petabytes. It has been used to sort 100 TB of data 3X faster than Hadoop MapReduce on 1/10th of the machines, winning the 2014 Daytona GraySort Benchmark, as well as to sort 1 PB. Several production workloads use Spark to do ETL and data analysis on PBs of data.

4)Does my data need to fit in memory to use Spark?

Answer)No. Spark's operators spill data to disk if it does not fit in memory, allowing it to run well on any sized data. Likewise, cached datasets that do not fit in memory are either spilled to disk or recomputed on the fly when needed, as determined by the RDD's storage level.

5)How can I run Spark on a cluster?

Answer)You can use either the standalone deploy mode, which only needs Java to be installed on each node, or the Mesos and YARN cluster managers. If you'd like to run on Amazon EC2, AMPLab provides EC2 scripts to automatically launch a cluster.

Note that you can also run Spark locally (possibly on multiple cores) without any special setup by just passing local[N] as the master URL, where N is the number of parallel threads you want.

6)Do I need Hadoop to run Spark?

Answer)No, but if you run on a cluster, you will need some form of shared file system (for example, NFS mounted at the same path on each node). If you have this type of filesystem, you can just deploy Spark in standalone mode.

7)Does Spark require modified versions of Scala or Python?

Answer)No. Spark requires no changes to Scala or compiler plugins. The Python API uses the standard CPython implementation, and can call into existing C libraries for Python such as NumPy.

8)We understand Spark Streaming uses micro-batching. Does this increase latency?

Answer)While Spark does use a micro-batch execution model, this does not have much impact on applications, because the batches can be as short as 0.5 seconds. In most applications of streaming big data, the analytics is done over a larger window (say 10 minutes), or the latency to get data in is higher (e.g. sensors collect readings every 10 seconds). Spark's model enables exactly-once semantics and consistency, meaning the system gives correct results despite slow nodes or failures.

9)Why Spark is good at low-latency iterative workloads e.g. Graphs and Machine Learning?

Answer)Machine Learning algorithms for instance logistic regression require many iterations before creating optimal resulting model. And similarly in graph algorithms which traverse all the nodes and edges. Any algorithm which needs many iteration before creating results can increase their performance when the intermediate partial results are stored in memory or at very fast solid state drives.

10)Which all kind of data processing supported by Spark?

Answer)Spark offers three kinds of data processing using batch, interactive (Spark Shell), and stream processing with the unified API and data structures.

www.smartdatacamp.com

11)Which all are the, ways to configure Spark Properties and order them least important to the most important?

Answer)Ans: There are the following ways to set up properties for Spark and user programs (in the order of importance from the least important to the most important):
conf/spark-defaults.conf : the default
--conf : the command line option used by spark-shell and spark-submit
SparkConf

12)What is the Default level of parallelism in Spark?

Answer)Default level of parallelism is the number of partitions when not specified explicitly by a user.

13) Is it possible to have multiple SparkContext in single JVM?

Answer)Yes, spark.driver.allowMultipleContexts is true (default: false). If true Spark logs warnings instead of throwing exceptions when multiple SparkContexts are active, i.e. multiple SparkContext are running in this JVM. When creating an instance of SparkContex.

14)What is the advantage of broadcasting values across Spark Cluster?

Answer)Spark transfers the value to Spark executors once, and tasks can share it without incurring repetitive network transmissions when requested multiple times.

15)How do you disable Info Message when running Spark Application

Answer)Under $SPARK_HOME/conf dir modify the log4j.properties file - change values INFO to ERROR

16)How do you evaluate your spark application for example i have access to a cluster (12 nodes where each node has 2 processors Intel(R) Xeon(R) CPU E5-2650 2.00GHz, where each processor has 8 cores), i want to know what are criteria that help me to tuning my application and to observe its performance.

Answer)For tuning you application you need to know few things

1) You Need to Monitor your application whether your cluster is under utilized or not how much resources are used by your application which you have created

Monitoring can be done using various tools eg. Ganglia From Ganglia you can find CPU, Memory and Network Usage.

2) Based on Observation about CPU and Memory Usage you can get a better idea what kind of tuning is needed for your application

Form Spark point of you

In spark-defaults.conf

you can specify what kind of serialization is needed how much Driver Memory and Executor Memory needed by your application even you can change Garbage collection algorithm.

Below are few Example you can tune this parameter based on your requirements
spark.serializer org.apache.spark.serializer.KryoSerializer
spark.driver.memory 5g
spark.executor.memory 3g
spark.executor.extraJavaOptions -XX:MaxPermSize=2G -XX:+UseG1GC
spark.driver.extraJavaOptions -XX:MaxPermSize=6G -XX:+UseG1GC

17)How do you define RDD?

Answer)A Resilient Distributed Dataset (RDD), the basic abstraction in Spark. It represents an immutable, partitioned collection of elements that can be operated on in parallel. Resilient Distributed Datasets (RDDs) are a distributed memory abstraction that lets programmers perform in-memory computations on large clusters in a fault-tolerant manner.

Resilient: Fault-tolerant and so able to recomputed missing or damaged partitions on node failures with the help of RDD lineage graph.

Distributed: across clusters.

Dataset: is a collection of partitioned data.

18)What is Lazy evaluated RDD mean?

Answer)Lazy evaluated, i.e. the data inside RDD is not available or transformed until an action is executed that triggers the execution.

19)How would you control the number of partitions of a RDD?

Answer)You can control the number of partitions of a RDD using repartition or coalesce operations.

www.smartdatacamp.com

20)Data is spread in all the nodes of cluster, how spark tries to process this data?

Answer)By default, Spark tries to read data into an RDD from the nodes that are close to it. Since Spark usually accesses distributed partitioned data, to optimize transformation operations it creates partitions to hold the data chunks

21)What is coalesce transformation?

Answer)The coalesce transformation is used to change the number of partitions. It can trigger RDD shuffling depending on the second shuffle boolean input parameter (defaults to false).

22)What is the difference between cache() and persist() method of RDD

Answer)RDDs can be cached (using RDD's cache() operation) or persisted (using RDD's persist(newLevel: StorageLevel) operation). The cache() operation is a synonym of persist() that uses the default storage level MEMORY_ONLY .

23)What is Shuffling?

Answer)Shuffling is a process of repartitioning (redistributing) data across partitions and may cause moving it across JVMs or even network when it is redistributed among executors.

Avoid shuffling at all cost. Think about ways to leverage existing partitions. Leverage partial aggregation to reduce data transfer.

24)What is the difference between groupByKey and use reduceByKey ?

Answer)Avoid groupByKey and use reduceByKey or combineByKey instead.
groupByKey shuffles all the data, which is slow.
reduceByKey shuffles only the results of sub-aggregations in each partition of the data.

25)What is checkpointing?

Answer)Checkpointing is a process of truncating RDD lineage graph and saving it to a reliable distributed (HDFS) or local file system. RDD checkpointing that saves the actual intermediate RDD data to a reliable distributed file system.

26)Define Spark architecture?

Answer) Spark uses a master/worker architecture. There is a driver that talks to a single coordinator called master that manages workers in which executors run. The driver and the executors run in their own Java processes

27)What are the workers?

Answer)Workers or slaves are running Spark instances where executors live to execute tasks. They are the compute nodes in Spark. A worker receives serialized/marshalled tasks that it runs in a thread pool.

28)Please explain, how worker's work, when a new Job submitted to them?

Answer)When SparkContext is created, each worker starts one executor. This is a separate java process or you can say new JVM, and it loads application jar in this JVM. Now executors connect back to your driver program and driver send them commands, like, foreach, filter, map etc. As soon as the driver quits, the executors shut down

29) Please define executors in detail?

Answer)Executors are distributed agents responsible for executing tasks. Executors provide in-memory storage for RDDs that are cached in Spark applications. When executors are started they register themselves with the driver and communicate directly to execute tasks.

30)What is DAGSchedular and how it performs?

Answer)DAGScheduler is the scheduling layer of Apache Spark that implements stage-oriented scheduling, i.e. after an RDD action has been called it becomes a job that is then transformed into a set of stages that are submitted as TaskSets for execution.

31)What is stage, with regards to Spark Job execution?

Answer)A stage is a set of parallel tasks, one per partition of an RDD, that compute partial results of a function executed as part of a Spark job.

32)What is Speculative Execution of a tasks?

www.smartdatacamp.com

Answer)Speculative tasks or task strugglers are tasks that run slower than most of the all tasks in a job.Speculative execution of tasks is a health-check procedure that checks for tasks to be speculated, i.e. running slower in a stage than the median of all successfully completed tasks in a taskset . Such slow tasks will be re-launched in another worker. It will not stop the slow tasks, but run a new copy in parallel.

33)Which all cluster manager can be used with Spark?

Answer)Apache Mesos, Hadoop YARN, Spark standalone

34)What is Data locality / placement?

Answer)Spark relies on data locality or data placement or proximity to data source, that makes Spark jobs sensitive to where the data is located. It is therefore important to have Spark running on Hadoop YARN cluster if the data comes from HDFS.

With HDFS the Spark driver contacts NameNode about the DataNodes (ideally local) containing the various blocks of a file or directory as well as their locations (represented as InputSplits), and then schedules the work to the SparkWorkers. Spark's compute nodes / workers should be running on storage nodes.

35)What is a Broadcast Variable?

Answer)Broadcast variables allow the programmer to keep a read-only variable cached on each machine rather than shipping a copy of it with tasks.

36)How can you define Spark Accumulators?

Answer)This are similar to counters in Hadoop MapReduce framework, which gives information regarding completion of tasks, or how much data is processed etc.

37)What is Apache Spark Streaming?

Answer)Spark Streaming helps to process live stream data. Data can be ingested from many sources like Kafka, Flume, Twitter, ZeroMQ, Kinesis, or TCP sockets, and can be processed using complex algorithms expressed with high-level functions like map, reduce, join and window.

38)Explain about transformations and actions in the context of RDDs?

Answer)Transformations are functions executed on demand, to produce a new RDD. All transformations are followed by actions. Some examples of transformations include map, filter and reduceByKey.

Actions are the results of RDD computations or transformations. After an action is performed, the data from RDD moves back to the local machine. Some examples of actions include reduce, collect, first, and take.

39)Can you use Spark to access and analyse data stored in Cassandra databases?

Answer)Yes, it is possible if you use Spark Cassandra Connector.

40)Is it possible to run Apache Spark on Apache Mesos?

Answer)Yes, Apache Spark can be run on the hardware clusters managed by Mesos.

41)How can you minimize data transfers when working with Spark?

Answer)Minimizing data transfers and avoiding shuffling helps write spark programs that run in a fast and reliable manner. The various ways in which data transfers can be minimized when working with Apache Spark are:

Using Broadcast Variable- Broadcast variable enhances the efficiency of joins between small and large RDDs.

Using Accumulators – Accumulators help update the values of variables in parallel while executing.

The most common way is to avoid operations ByKey, repartition or any other operations which trigger shuffles.

42)What is the significance of Sliding Window operation?

Answer)Sliding Window controls transmission of data packets between various computer networks. Spark Streaming library provides windowed computations where the transformations on RDDs are applied over a sliding window of data. Whenever the window slides, the RDDs that fall within the particular window are combined and operated upon to produce new RDDs of the windowed DStream.

43)What is a DStream?

www.smartdatacamp.com

Answer)Discretized Stream is a sequence of Resilient Distributed Databases that represent a stream of data. DStreams can be created from various sources like Apache Kafka, HDFS, and Apache Flume. DStreams have two operations –

Transformations that produce a new DStream.

Output operations that write data to an external system.

44)Which one will you choose for a project –Hadoop MapReduce or Apache Spark?

Answer)The answer to this question depends on the given project scenario - as it is known that Spark makes use of memory instead of network and disk I/O. However, Spark uses large amount of RAM and requires dedicated machine to produce effective results. So the decision to use Hadoop or Spark varies dynamically with the requirements of the project and budget of the organization

45)What are the various levels of persistence in Apache Spark?

Answer)Apache Spark automatically persists the intermediary data from various shuffle operations, however it is often suggested that users call persist () method on the RDD in case they plan to reuse it. Spark has various persistence levels to store the RDDs on disk or in memory or as a combination of both with different replication levels.

The various storage/persistence levels in Spark are -
MEMORY_ONLY
MEMORY_ONLY_SER
MEMORY_AND_DISK
MEMORY_AND_DISK_SER,
DISK_ONLY
OFF_HEAP

46)Explain the difference between Spark SQL and Hive?

Answer)Spark SQL is faster than Hive.

Any Hive query can easily be executed in Spark SQL but vice-versa is not true.

Spark SQL is a library whereas Hive is a framework.

It is not mandatory to create a metastore in Spark SQL but it is mandatory to create a Hive metastore.

Spark SQL automatically infers the schema whereas in Hive schema needs to be explicitly declared.

47)What does a Spark Engine do?

Answer)Spark Engine is responsible for scheduling, distributing and monitoring the data application across the cluster.

48)What are benefits of Spark over MapReduce?

Answer)Due to the availability of in-memory processing, Spark implements the processing around 10-100x faster than Hadoop MapReduce. MapReduce makes use of persistence storage for any of the data processing tasks.

Unlike Hadoop, Spark provides in-built libraries to perform multiple tasks form the same core like batch processing, Steaming, Machine learning, Interactive SQL queries. However, Hadoop only supports batch processing.

Hadoop is highly disk-dependent whereas Spark promotes caching and in-memory data storage.

Spark is capable of performing computations multiple times on the same dataset. This is called iterative computation while there is no iterative computing implemented by Hadoop.

49)What is Spark Driver?

Answer)Spark Driver is the program that runs on the master node of the machine and declares transformations and actions on data RDDs. In simple terms, a driver in Spark creates SparkContext, connected to a given Spark Master.

The driver also delivers the RDD graphs to Master, where the standalone cluster manager runs.

50)What is DataFrames?

Answer)It is a collection of data which organize in named columns. It is theoretically equivalent to a table in relational database. But it is more optimized. Just like RDD, DataFrames evaluates lazily. Using lazy evaluation we can optimize the execution. It optimizes by applying the techniques such as bytecode generation and predicate push-downs

51)What are the advantages of DataFrame?

Answer)It makes large data set processing even easier. Data Frame also allows developers to impose a structure onto a distributed collection of data. As a result, it allows higher-level abstraction.

Data frame is both space and performance efficient.

www.smartdatacamp.com

It can deal with both structured and unstructured data formats, for example, Avro, CSV etc . And also storage systems like HDFS, HIVE tables, MySQL, etc.

The DataFrame API's are available in various programming languages. For example Java, Scala, Python, and R.

It provides Hive compatibility. As a result, we can run unmodified Hive queries on existing Hive warehouse.

Catalyst tree transformation uses DataFrame in four phases: a) Analyze logical plan to solve references. b) Logical plan optimization c) Physical planning d) Code generation to compile part of the query to Java bytecode.

It can scale from kilobytes of data on the single laptop to petabytes of data on the large cluster.

52)What is write ahead log(journaling)?

Answer)The write-ahead log is a technique that provides durability in a database system. It works in the way that all the operation that applies on data, we write it to write-ahead log. The logs are durable in nature. Thus, when the failure occurs we can easily recover the data from these logs. When we enable the write-ahead log Spark stores the data in fault-tolerant file system.

53)How Spark Streaming API works?

Answer)Programmer set a specific time in the configuration, with in this time how much data gets into the Spark, that data separates as a batch. The input stream (DStream) goes into spark streaming. Framework breaks up into small chunks called batches, then feeds into the spark engine for processing. Spark Streaming API passes that batches to the core engine. Core engine can generate the final results in the form of streaming batches. The output also in the form of batches. It can allows streaming data and batch data for processing.

54)If there is certain data that we want to use again and again in different transformations what should improve the performance?

Answer)RDD can be persisted or cached. There are various ways in which it can be persisted: in-memory, on disc etc. So, if there is a dataset that needs a good amount computing to arrive at, you should consider caching it. You can cache it to disc if preparing it again is far costlier than just reading from disc or it is very huge in size and would not fit in the RAM. You can cache it to memory if it can fit into the memory.

55)What happens to RDD when one of the nodes on which it is distributed goes down?

Answer)Since Spark knows how to prepare a certain data set because it is aware of various transformations and actions that have lead to the dataset, it will be able to apply the same transformations and actions to prepare the lost partition of the node which has gone down.

56)How do I skip a header from CSV files in Spark?

Answer)Spark 2.x : spark.read.format("csv").option("header","true").load("filePath")

57)Have you ever encounter Spark java.lang.OutOfMemoryError? How to fix this issue?

Answer)I have a few suggestions:

If your nodes are configured to have 6g maximum for Spark (and are leaving a little for other processes), then use 6g rather than 4g, spark.executor.memory=6g. Make sure you're using as much memory as possible by checking the UI (it will say how much mem you're using)

Try using more partitions, you should have 2 - 4 per CPU. IME increasing the number of partitions is often the easiest way to make a program more stable (and often faster). For huge amounts of data you may need way more than 4 per CPU, I've had to use 8000 partitions in some cases!

Decrease the fraction of memory reserved for caching, using spark.storage.memoryFraction. If you don't use cache() or persist in your code, this might as well be 0. It's default is 0.6, which means you only get 0.4 * 4g memory for your heap. IME reducing the mem frac often makes OOMs go away. UPDATE: From spark 1.6 apparently we will no longer need to play with these values, spark will determine them automatically.

Similar to above but shuffle memory fraction. If your job doesn't need much shuffle memory then set it to a lower value (this might cause your shuffles to spill to disk which can have catastrophic impact on speed). Sometimes when it's a shuffle operation that's OOMing you need to do the opposite i.e. set it to something large, like 0.8, or make sure you allow your shuffles to spill to disk (it's the default since 1.0.0).

Watch out for memory leaks, these are often caused by accidentally closing over objects you don't need in your lambdas. The way to diagnose is to look out for the "task serialized as XXX bytes" in the logs, if XXX is larger than a few k or more than an MB, you may have a memory leak.

Related to above; use broadcast variables if you really do need large objects.

If you are caching large RDDs and can sacrifice some access time consider serialising the RDD Or even caching them on disk (which sometimes isn't that bad if using SSDs).

58)Does SparkSQL support subquery?

Answer)Spark 2.0+

Spark SQL should support both correlated and uncorrelated subqueries. See SubquerySuite for details. Some examples include:

select * from l where exists (select * from r where l.a = r.c)
select * from l where not exists (select * from r where l.a = r.c)
select * from l where l.a in (select c from r)
select * from l where a not in (select c from r)

Unfortunately as for now (Spark 2.0) it is impossible to express the same logic using DataFrame DSL.
Spark < 2.0
Spark supports subqueries in the FROM clause (same as Hive <= 0.12).
SELECT col FROM (SELECT * FROM t1 WHERE bar) t2

59)How to read multiple text files into a single RDD?

Answer)sc.textFile("/my/dir1,/my/paths/part-00[0-5]*,/another/dir,/a/specific/file")

60)What is the difference between map and flatMap and a good use case for each?

Answer)Generally we use word count example in hadoop. I will take the same use case and will use map and flatMap and we will see the difference how it is processing the data.

Below is the sample data file.
hadoop is fast
hive is sql on hdfs
spark is superfast
spark is awesome
The above file will be parsed using map and flatMap.
Using map
wc = data.map(lambda line:line.split(" "));
wc.collect()
[u'hadoop is fast', u'hive is sql on hdfs', u'spark is superfast', u'spark is awesome']
Input has 4 lines and output size is 4 as well, i.e., N elements ==> N elements.
Using flatMap
fm = data.flatMap(lambda line:line.split(" "));
fm.collect()
[u'hadoop', u'is', u'fast', u'hive', u'is', u'sql', u'on', u'hdfs', u'spark', u'is', u'superfast', u'spark', u'is', u'awesome']

61)What are the various data sources available in SparkSQL?

Answer)CSV file, Parquet file, JSON Datasets, Hive Table.

62)What are the key features of Apache Spark that you like?

Answer)Spark provides advanced analytic options like graph algorithms, machine learning, streaming data, etc.It has built-in APIs in multiple languages like Java, Scala, Python and R.It has good performance gains, as it helps run an application in the Hadoop cluster ten times faster on disk and 100 times faster in memory.

63)Name some sources from where Spark streaming component can process realtime data.

Answer)Apache Flume, Apache Kafka, Amazon Kinesis

64)Name some companies that are already using Spark Streaming.

Answer)Uber, Netflix, Pinterest.

65)What do you understand by receivers in Spark Streaming ?

Answer)Receivers are special entities in Spark Streaming that consume data from various data sources and move them to Apache Spark. Receivers are usually created by streaming contexts as long running tasks on various executors and scheduled to operate in a round robin manner with each receiver taking a single core.

66)What is GraphX?

Answer)Spark uses GraphX for graph processing to build and transform interactive graphs. The GraphX component enables programmers to reason about structured data at scale.

67)What does MLlib do?

Answer)MLlib is scalable machine learning library provided by Spark. It aims at making machine learning easy and scalable with common learning algorithms and use cases like clustering, regression filtering, dimensional reduction, and alike.

68)What is PageRank?

www.smartdatacamp.com

Answer)A unique feature and algorithm in graph, PageRank is the measure of each vertex in the graph. For instance, an edge from u to v represents endorsement of vâ€™s importance by u. In simple terms, if a user at Instagram is followed massively, it will rank high on that platform.

69)Do you need to install Spark on all nodes of Yarn cluster while running Spark on Yarn?

Answer)No because Spark runs on top of Yarn.

70)List the advantage of Parquet file in Apache Spark.

Answer)Parquet is a columnar format supported by many data processing systems. The benefits of having a columnar storage are
1)Columnar storage limits IO operations.
2)Columnar storage can fetch specific columns that you need to access.
3)Columnar storage consumes less space.
4)Columnar storage gives better-summarized data and follows type-specific encoding.

Apache Kafka

Apache Kafka is an open-source stream processing platform developed by the Apache Software Foundation written in Scala and Java. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds.

1)What are the advantages of using Apache Kafka?

Answer) The Advantages of using Apache Kafka are as follows:

High Throughput:The design of Kafka enables the platform to process messages at very fast speed. The processing rates in Kafka can exceed beyond 100k/seconds. The data is processed in a partitioned and ordered fashion.

Scalability:The scalability can be achieved in Kafka at various levels. Multiple producers can write to the same topic. Topics can be partitioned. Consumers can be grouped to consume individual partitions.

Fault Tolerance:Kafka is a distributed architecture which means there are several nodes running together to serve the cluster. Topics inside Kafka are replicated. Users can choose the number of replicas for each topic to be safe in case of a node failure. Node failure in cluster won't impact. Integration with Zookeeper provides producers and consumers accurate information about the cluster. Internally each topic has its own leader which takes care of the writes. Failure of node ensures new leader election.

Durability:Kafka offers data durability as well. The message written in Kafka can be persisted. The persistence can be configured. This ensures re-processing, if required, can be performed.

2)Which are the elements of Kafka?

Answer)The most important elements of Kafka:
Topic – It is the bunch of similar kind of messages
Producer – using this one can issue communications to the topic
Consumer – it endures to a variety of topics and takes data from brokers.
Brokers – this is the place where the issued messages are stored

3)What is Kafka Logs?

Answer)An important concept for Apache Kafka is "log". This is not related to application log or system log. This is a log of the data. It creates a loose structure of the data which is consumed by Kafka. The notion of "log" is an ordered, append-only sequence of data. The data can be anything because for Kafka it will be just an array of bytes.

www.smartdatacamp.com

4)Explain What Is Zookeeper In Kafka? Can We Use Kafka Without Zookeeper?

Answer)Zookeeper is an open source, high-performance co-ordination service used for distributed applications adapted by Kafka.

No, it is not possible to bye-pass Zookeeper and connect straight to the Kafka broker. Once the Zookeeper is down, it cannot serve client request.

Zookeeper is basically used to communicate between different nodes in a cluster

In Kafka, it is used to commit offset, so if node fails in any case it can be retrieved from the previously committed offset

Apart from this it also does other activities like leader detection, distributed synchronization, configuration management, identifies when a new node leaves or joins, the cluster, node status in real time, etc.

5)Why Replication Is Required In Kafka?

Answer)Replication of message in Kafka ensures that any published message does not lose and can be consumed in case of machine error, program error or more common software upgrades.

6)What is the role of offset in Kafka?

Answer)Offset is nothing but an unique id that is assigned to the partitions. The messages are contained in this partitions. The important aspect or use of offset is that it identifies every message with the id which is available within the partition.

7)What is a consumer group?

Answer)A consumer group is nothing but an exclusive concept of Kafka.

Within each and every Kafka consumer group, we will have one or more consumers who actually consume subscribed topics.

8)What is the core API in Kafka?

Answer)They are four main core API's:
1. Producer API
2. Consumer API
3. Streams API

4. Connector API

9)Explain the functionality of producer API in Kafka?

Answer)The producer API is responsible where it will allow the application to push a stream of records to one of the Kafka topics.

10)Explain the functionality of Consumer API in Kafka?

Answer)The Consumer API is responsible where it allows the application to receive one or more topics and at the same time process the stream of data that is produced.

11)Explain the functionality of Streams API in Kafka?

Answer)The Streams API is responsible where it allows the application to act as a processor and within the process, it will be effectively transforming the input streams to output streams.

12)Explain the functionality of Connector API in Kafka?

Answer)The Connector API is responsible where it allows the application to stay connected and keeping a track of all the changes that happen within the system. For this to happen, we will be using reusable producers and consumers which stays connected to the Kafka topics.

13)Explain what is a topic?

Answer)A topic is nothing but a category classification or it can be a feed name out of which the records are actually published. Topics are always classified, the multi subscriber.

14)What is the purpose of retention period in Kafka cluster?

Answer)Within the Kafka cluster, it retains all the published records. It doesn't check whether they have been consumed or not. Using a configuration setting for the retention period, the records can be discarded. The main reason to discard the records from the Kafka cluster is that it can free up some space.

15)Mention what is the maximum size of the message does Kafka server can receive?

www.smartdatacamp.com

Answer)The maximum size of the message that Kafka server can receive is 1000000 bytes.

16)Explain how you can improve the throughput of a remote consumer?

Answer)If the consumer is located in a different data center from the broker, you may require to tune the socket buffer size to amortize the long network latency.

17)Is it possible to get the message offset after producing?

Answer)You cannot do that from a class that behaves as a producer like in most queue systems, its role is to fire and forget the messages. The broker will do the rest of the work like appropriate metadata handling with id's, offsets, etc.

As a consumer of the message, you can get the offset from a Kafka broker. If you gaze in the SimpleConsumer class, you will notice it fetches MultiFetchResponse objects that include offsets as a list. In addition to that, when you iterate the Kafka Message, you will have MessageAndOffset objects that include both, the offset and the message sent.

18)Explain the concept of Leader and Follower?

Answer)Every partition in Kafka has one server which plays the role of a Leader, and none or more servers that act as Followers. The Leader performs the task of all read and write requests for the partition, while the role of the Followers is to passively replicate the leader. In the event of the Leader failing, one of the Followers will take on the role of the Leader. This ensures load balancing of the server.

19)If a Replica stays out of the ISR for a long time, what does it signify?

Answer)It means that the Follower is unable to fetch data as fast as data accumulated by the Leader.

20)How do you define a Partitioning Key?

Answer)Within the Producer, the role of a Partitioning Key is to indicate the destination partition of the message. By default, a hashing-based Partitioner is used to determine the partition ID given the key. Alternatively, users can also use customized Partitions.

21)In the Producer, when does QueueFullException occur?

Answer)QueueFullException typically occurs when the Producer attempts to send messages at a pace that the Broker cannot handle. Since the Producer doesn't block, users will need to add enough brokers to collaboratively handle the increased load.

22)Kafka Stream application failed to start, with the a rocksDB exception raised as "java.lang.ExceptionInInitializerError.. Unable to load the RocksDB shared libraryjava". How to resolve this?

Answer)Streams API uses RocksDB as the default local persistent key-value store. And RocksDB JNI would try to statically load the sharedlibs into java.io.tmpdir. On Unix-like platforms, the default value of this system environment property is typically /tmp, or /var/tmp; On Microsoft Windows systems the property is typically C:\\WINNT\\TEMP.

If your application does not have permission to access these directories (or for Unix-like platforms if the pointed location is not mounted), the above error would be thrown. To fix this, you can either grant the permission to access this directory to your applications, or change this property when executing your application like java -Djava.io.tmpdir=[].

23)Have you encountered Kafka Stream application's memory usage keeps increasing when running until it hits an OOM. Why any specific reason?

Answer)The most common cause of this scenario is that you did not close an iterator from the state stores after completed using it. For persistent stores like RocksDB, an iterator is usually backed by some physical resources like open file handlers and in-memory caches. Not closing these iterators would effectively causing resource leaks.

24)Extracted timestamp value is negative, which is not allowed. What does this mean in Kafka Streaming?

Answer)This error means that the timestamp extractor of your Kafka Streams application failed to extract a valid timestamp from a record. Typically, this points to a problem with the record (e.g., the record does not contain a timestamp at all), but it could also indicate a problem or bug in the timestamp extractor used by the application.

25)How to scale a Streams app,that is increase number of input partitions?

Answer)Basically, Kafka Streams does not allow to change the number of input topic partitions during its life time. If you stop a running Kafka Streams application, change the number of input

topic partitions, and restart your app it will most likely break with an exception as described in FAQ "What does exception "Store someStoreName's change log (someStoreName-changelog) does not contain partition someNumber mean? It is tricky to fix this for production use cases and it is highly recommended to not change the number of input topic partitions (cf. comment below). For POC/demos it's not difficult to fix though.

In order to fix this, you should reset your application using Kafka's application reset tool: Kafka Streams Application Reset Tool.

26)I get a locking exception similar to "Caused by: java.io.IOException: Failed to lock the state directory: /tmp/kafka-streams/app-id/0_0". How can I resolve this?

Answer)You can apply the following workaround:

switch to single threaded execution

if you want to scale your app, start multiple instances (instead of going multi-threaded with one instance)

if you start multiple instances on the same host, use a different state directory (state.dir config parameter) for each instance (to "isolate" the instances from each other)

It might also be necessary, to delete the state directory manually before starting the application. This will not result in data loss – the state will be recreated from the underlying changelog topic.0

27)How should I set metadata.broker.list?

Answer)The broker list provided to the producer is only used for fetching metadata. Once the metadata response is received, the producer will send produce requests to the broker hosting the corresponding topic/partition directly, using the ip/port the broker registered in ZK. Any broker can serve metadata requests. The client is responsible for making sure that at least one of the brokers in metadata.broker.list is accessible. One way to achieve this is to use a VIP in a load balancer. If brokers change in a cluster, one can just update the hosts associated with the VIP.

28)Why do I get QueueFullException in my producer when running in async mode?

Answer)This typically happens when the producer is trying to send messages quicker than the broker can handle. If the producer can't block, one will have to add enough brokers so that they jointly can handle the load. If the producer can block, one can set queue.enqueueTimeout.ms in producer config to -1. This way, if the queue is full, the producer will block instead of dropping messages.

29)Why are my brokers not receiving producer sent messages?

Answer)This happened when I tried to enable gzip compression by setting compression.codec to 1. With the code change, not a single message was received by the brokers even though I had called producer.send() 1 million times. No error printed by producer and no error could be found in broker's kafka-request.log. By adding log4j.properties to my producer's classpath and switching the log level to DEBUG, I captured the java.lang.NoClassDefFoundError: org/xerial/snappy/SnappyInputStream thrown at the producer side. Now I can see this error can be resolved by adding snappy jar to the producer's classpath.

30)Is it possible to delete a topic?

Answer)Deleting a topic is supported since 0.8.2.x. You will need to enable topic deletion (setting delete.topic.enable to true) on all brokers first.

31)Why does Kafka consumer never get any data?

Answer)By default, when a consumer is started for the very first time, it ignores all existing data in a topic and will only consume new data coming in after the consumer is started. If this is the case, try sending some more data after the consumer is started. Alternatively, you can configure the consumer by setting auto.offset.reset to earliest for the new consumer in 0.9 and smallest for the old consumer.

32)Why does Kafka consumer get InvalidMessageSizeException?

Answer)This typically means that the fetch size of the consumer is too small. Each time the consumer pulls data from the broker, it reads bytes up to a configured limit. If that limit is smaller than the largest single message stored in Kafka, the consumer can't decode the message properly and will throw an InvalidMessageSizeException. To fix this, increase the limit by setting the property fetch.size (0.7) / fetch.message.max.bytes (0.8) properly in config/consumer.properties. The default fetch.size is 300,000 bytes. For the new consumer in 0.9, the property to adjust is max.partition.fetch.bytes, and the default is 1MB.

33)Should I choose multiple group ids or a single one for the consumers?

Answer)If all consumers use the same group id, messages in a topic are distributed among those consumers. In other words, each consumer will get a non-overlapping subset of the messages. Having more consumers in the same group increases the degree of parallelism and the overall throughput of consumption. See the next question for the choice of the number of consumer instances. On the other hand, if each consumer is in its own group, each consumer will get a full copy of all messages.

34)Why some of the consumers in a consumer group never receive any message?

Answer)Currently, a topic partition is the smallest unit that we distribute messages among consumers in the same consumer group. So, if the number of consumers is larger than the total number of partitions in a Kafka cluster (across all brokers), some consumers will never get any data. The solution is to increase the number of partitions on the broker.

35)Why are there many rebalances in Kafka consumer log?

Answer)A typical reason for many rebalances is the consumer side GC. If so, you will see Zookeeper session expirations in the consumer log (grep for Expired). Occasional rebalances are fine. Too many rebalances can slow down the consumption and one will need to tune the java GC setting.

36)Why messages are delayed in kafka consumer?

Answer)This could be a general throughput issue. If so, you can use more consumer streams (may need to increase # partitions) or make the consumption logic more efficient.

Another potential issue is when multiple topics are consumed in the same consumer connector. Internally, we have an in-memory queue for each topic, which feed the consumer iterators. We have a single fetcher thread per broker that issues multi-fetch requests for all topics. The fetcher thread iterates the fetched data and tries to put the data for different topics into its own in-memory queue. If one of the consumer is slow, eventually its corresponding in-memory queue will be full. As a result, the fetcher thread will block on putting data into that queue. Until that queue has more space, no data will be put into the queue for other topics. Therefore, those other topics, even if they have less volume, their consumption will be delayed because of that. To address this issue, either making sure that all consumers can keep up, or using separate consumer connectors for different topics.

37)How to improve the throughput of a remote consumer?

Answer)If the consumer is in a different data center from the broker, you may need to tune the socket buffer size to amortize the long network latency. Specifically, for Kafka 0.7, you can increase socket.receive.buffer in the broker, and socket.buffersize and fetch.size in the consumer. For Kafka 0.8, the consumer properties are socket.receive.buffer.bytes and fetch.message.max.bytes.

38)How to consume large messages?

Answer)First you need to make sure these large messages can be accepted at Kafka brokers. The broker property message.max.bytes controls the maximum size of a message that can be accepted at the broker, and any single message (including the wrapper message for compressed message set) whose size is larger than this value will be rejected for producing. Then you need to make sure consumers can fetch such large messages from brokers. For the old consumer, you should use the property fetch.message.max.bytes, which controls the maximum number of bytes a consumer issues in one fetch. If it is less than a message's size, the fetching will be blocked on that message keep retrying. The property for the new consumer is max.partition.fetch.bytes.

39)How does Kafka depend on Zookeeper?

Answer)Starting from 0.9, we are removing all the Zookeeper dependency from the clients (for details one can check this page). However, the brokers will continue to be heavily depend on Zookeeper for:
Server failure detection.
Data partitioning.
In-sync data replication.

Once the Zookeeper quorum is down, brokers could result in a bad state and could not normally serve client requests, etc. Although when Zookeeper quorum recovers, the Kafka brokers should be able to resume to normal state automatically, there are still a few corner cases the they cannot and a hard kill-and-recovery is required to bring it back to normal. Hence it is recommended to closely monitor your zookeeper cluster and provision it so that it is performant.

Also note that if Zookeeper was hard killed previously, upon restart it may not successfully load all the data and update their creation timestamp. To resolve this you can clean-up the data directory of the Zookeeper before restarting (if you have critical metadata such as consumer offsets you would need to export / import them before / after you cleanup the Zookeeper data and restart the server).

40)Why can't Kafka consumers/producers connect to the brokers? What could be the reason?

Answer)When a broker starts up, it registers its ip/port in ZK. You need to make sure the registered ip is consistent with what's listed in metadata.broker.list in the producer config. By default, the registered ip is given by InetAddress.getLocalHost.getHostAddress. Typically, this should return the real ip of the host. However, sometimes (e.g., in EC2), the returned ip is an internal one and can't be connected to from outside. The solution is to explicitly set the host ip to be registered in ZK by setting the "hostname" property in server.properties. In another rare case where the binding host/port is different from the host/port for client connection, you can set advertised.host.name and advertised.port for client connection.

41)How many topics can I have?

www.smartdatacamp.com

Answer)Unlike many messaging systems Kafka topics are meant to scale up arbitrarily. Hence we encourage fewer large topics rather than many small topics. So for example if we were storing notifications for users we would encourage a design with a single notifications topic partitioned by user id rather than a separate topic per user.

The actual scalability is for the most part determined by the number of total partitions across all topics not the number of topics itself (see the question below for details).

42)How do I choose the number of partitions for a topic?

Answer)There isn't really a right answer, we expose this as an option because it is a tradeoff. The simple answer is that the partition count determines the maximum consumer parallelism and so you should set a partition count based on the maximum consumer parallelism you would expect to need (i.e. over-provision). Clusters with up to 10k total partitions are quite workable. Beyond that we don't aggressively test (it should work, but we can't guarantee it).

Here is a more complete list of tradeoffs to consider:

A partition is basically a directory of log files.

Each partition must fit entirely on one machine. So if you have only one partition in your topic you cannot scale your write rate or retention beyond the capability of a single machine. If you have 1000 partitions you could potentially use 1000 machines.

Each partition is totally ordered. If you want a total order over all writes you probably want to have just one partition.

Each partition is not consumed by more than one consumer thread/process in each consumer group. This allows to have each process consume in a single threaded fashion to guarantee ordering to the consumer within the partition (if we split up a partition of ordered messages and handed them out to multiple consumers even though the messages were stored in order they would be processed out of order at times).

Many partitions can be consumed by a single process, though. So you can have 1000 partitions all consumed by a single process.

Another way to say the above is that the partition count is a bound on the maximum consumer parallelism.

More partitions will mean more files and hence can lead to smaller writes if you don't have enough memory to properly buffer the writes and coalesce them into larger writes

Each partition corresponds to several znodes in zookeeper. Zookeeper keeps everything in memory so this can eventually get out of hand.

More partitions means longer leader fail-over time. Each partition can be handled quickly (milliseconds) but with thousands of partitions this can add up.

When we checkpoint the consumer position we store one offset per partition so the more partitions the more expensive the position checkpoint is.

It is possible to later expand the number of partitions BUT when we do so we do not attempt to reorganize the data in the topic. So if you are depending on key-based semantic partitioning in

your processing you will have to manually copy data from the old low partition topic to a new higher partition topic if you later need to expand.

43)How to replace a failed broker?

Answer)When a broker fails, Kafka doesn't automatically re-replicate the data on the failed broker to other brokers. This is because in the common case, one brings down a broker to apply code or config changes, and will bring up the broker quickly afterward. Re-replicating the data in this case will be wasteful. In the rarer case that a broker fails completely, one will need to bring up another broker with the same broker id on a new server. The new broker will automatically replicate the missing data.

44)Can I add new brokers dynamically to a cluster?

Answer)Yes, new brokers can be added online to a cluster. Those new brokers won't have any data initially until either some new topics are created or some replicas are moved to them using the partition reassignment tool.

www.smartdatacamp.com

Apache Sqoop

Apache Sqoop is a command-line interface application for transferring data between relational databases and Hadoop.

1)What is the default file format to import data using Apache Sqoop?

Answer)Sqoop allows data to be imported using two file formats
i) Delimited Text File Format
This is the default file format to import data using Sqoop. This file format can be explicitly specified using the –as-textfile argument to the import command in Sqoop. Passing this as an argument to the command will produce the string based representation of all the records to the output files with the delimited characters between rows and columns.
ii) Sequence File Format
It is a binary file format where records are stored in custom record-specific data types which are shown as Java classes. Sqoop automatically creates these data types and manifests them as java classes.

2)How do I resolve a Communications Link Failure when connecting to MySQL?

Answer)Verify that you can connect to the database from the node where you are running Sqoop:
$ mysql --host=IP Address --database=test --user=username --password=password
Add the network port for the server to your my.cnf file
Set up a user account to connect via Sqoop. Grant permissions to the user to access the database over the network:
Log into MySQL as root mysql -u root -p ThisIsMyPassword
Issue the following command: mysql> grant all privileges on test.* to 'testuser'@'%' identified by 'testpassword'

3)How do I resolve an IllegalArgumentException when connecting to Oracle?

Answer)This could be caused a non-owner trying to connect to the table so prefix the table name with the schema, for example SchemaName.OracleTableName.

4)What's causing this Exception in thread main java.lang.IncompatibleClassChangeError when running non-CDH Hadoop with Sqoop?

Answer)Try building Sqoop 1.4.1-incubating with the command line property -Dhadoopversion=20.

5)How do I resolve an ORA-00933 error SQL command not properly ended when connecting to Oracle?

Answer)Omit the option --driver oracle.jdbc.driver.OracleDriver and then re-run the Sqoop command.

6)I have around 300 tables in a database. I want to import all the tables from the database except the tables named Table298, Table 123, and Table299. How can I do this without having to import the tables one by one?

Answer)This can be accomplished using the import-all-tables import command in Sqoop and by specifying the exclude-tables option with it as follows-
sqoop import-all-tables
--connect –username –password --exclude-tables Table298, Table 123, Table 299

7)Does Apache Sqoop have a default database?

Answer)Yes, MySQL is the default database.

8)How can I import large objects (BLOB and CLOB objects) in Apache Sqoop?

Answer)Apache Sqoop import command does not support direct import of BLOB and CLOB large objects. To import large objects, I Sqoop, JDBC based imports have to be used without the direct argument to the import utility.

9)How can you execute a free form SQL query in Sqoop to import the rows in a sequential manner?

Answer)This can be accomplished using the –m 1 option in the Sqoop import command. It will create only one MapReduce task which will then import rows serially.

10)How will you list all the columns of a table using Apache Sqoop?

Answer)Unlike sqoop-list-tables and sqoop-list-databases, there is no direct command like sqoop-list-columns to list all the columns. The indirect way of achieving this is to retrieve the columns of the desired tables and redirect them to a file which can be viewed manually containing the column names of a particular table.

www.smartdatacamp.com

sqoop import --m 1 --connect jdbc: sqlserver: nameofmyserver; database=nameofmydatabase; username=DeZyre; password=mypassword --query SELECT column_name, DATA_TYPE FROM INFORMATION_SCHEMA.Columns WHERE table_name=mytableofinterest AND $CONDITIONS --target-dir mytableofinterest_column_name

11)What is the difference between Sqoop and DistCP command in Hadoop?

Answer)Both distCP (Distributed Copy in Hadoop) and Sqoop transfer data in parallel but the only difference is that distCP command can transfer any kind of data from one Hadoop cluster to another whereas Sqoop transfers data between RDBMS and other components in the Hadoop ecosystem like HBase, Hive, HDFS, etc.

12)What is Sqoop metastore?

Answer)Sqoop metastore is a shared metadata repository for remote users to define and execute saved jobs created using sqoop job defined in the metastore. The sqoop –site.xml should be configured to connect to the metastore.

13)What is the significance of using –split-by clause for running parallel import tasks in Apache Sqoop?

Answer)--Split-by clause is used to specify the columns of the table that are used to generate splits for data imports. This clause specifies the columns that will be used for splitting when importing the data into the Hadoop cluster. —split-by clause helps achieve improved performance through greater parallelism. Apache Sqoop will create splits based on the values present in the columns specified in the –split-by clause of the import command. If the –split-by clause is not specified, then the primary key of the table is used to create the splits while data import. At times the primary key of the table might not have evenly distributed values between the minimum and maximum range. Under such circumstances –split-by clause can be used to specify some other column that has even distribution of data to create splits so that data import is efficient.

14)You use –split-by clause but it still does not give optimal performance then how will you improve the performance further.

Answer)Using the –boundary-query clause. Generally, sqoop uses the SQL query select min (), max () from to find out the boundary values for creating splits. However, if this query is not optimal then using the –boundary-query argument any random query can be written to generate two numeric columns.

15)During sqoop import, you use the clause –m or –numb-mappers to specify the number of mappers as 8 so that it can run eight parallel MapReduce tasks, however, sqoop runs only four parallel MapReduce tasks. Why?

Answer)Hadoop MapReduce cluster is configured to run a maximum of 4 parallel MapReduce tasks and the sqoop import can be configured with number of parallel tasks less than or equal to 4 but not more than 4.

16)You successfully imported a table using Apache Sqoop to HBase but when you query the table it is found that the number of rows is less than expected. What could be the likely reason?

Answer)If the imported records have rows that contain null values for all the columns, then probably those records might have been dropped off during import because HBase does not allow null values in all the columns of a record.

17)The incoming value from HDFS for a particular column is NULL. How will you load that row into RDBMS in which the columns are defined as NOT NULL?

Answer)Using the –input-null-string parameter, a default value can be specified so that the row gets inserted with the default value for the column that it has a NULL value in HDFS.

18)If the source data gets updated every now and then, how will you synchronise the data in HDFS that is imported by Sqoop?

Answer)Data can be synchronised using incremental parameter with data import –
--Incremental parameter can be used with one of the two options-
i) append-If the table is getting updated continuously with new rows and increasing row id values then incremental import with append option should be used where values of some of the columns are checked (columns to be checked are specified using –check-column) and if it discovers any modified value for those columns then only a new row will be inserted.
ii) lastmodified – In this kind of incremental import, the source has a date column which is checked for. Any records that have been updated after the last import based on the lastmodifed column in the source, the values would be updated.

19)Below command is used to specify the connect string that contains hostname to connect MySQL with local host and database name as test_db
––connect jdbc: mysql: //localhost/test_db
Is the above command the best way to specify the connect string in case I want to use Apache Sqoop with a distributed hadoop cluster?

www.smartdatacamp.com

Answer)When using Sqoop with a distributed Hadoop cluster the URL should not be specified with localhost in the connect string because the connect string will be applied on all the DataNodes with the Hadoop cluster. So, if the literal name localhost is mentioned instead of the IP address or the complete hostname then each node will connect to a different database on their localhosts. It is always suggested to specify the hostname that can be seen by all remote nodes.

20)What are the relational databases supported in Sqoop?

Answer)Below are the list of RDBMSs that are supported by Sqoop Currently.
MySQL
PostGreSQL
Oracle
Microsoft SQL
IBM's Netezza
Teradata

21)What are the destination types allowed in Sqoop Import command?

Answer)Currently Sqoop Supports data imported into below services.
HDFS
Hive
HBase
HCatalog
Accumulo

22)Is Sqoop similar to distcp in hadoop?

Answer)Partially yes, hadoop's distcp command is similar to Sqoop Import command. Both submits parallel map-only jobs but distcp is used to copy any type of files from Local FS/HDFS to HDFS and Sqoop is for transferring the data records only between RDMBS and Hadoop eco system services, HDFS, Hive and HBase.

23)What are the majorly used commands in Sqoop?

Answer)In Sqoop Majorly Import and export commands are used. But below commands are also useful some times.
codegen
eval
import-all-tables
job
list-databases

list-tables
merge
metastore

24)While loading tables from MySQL into HDFS, if we need to copy tables with maximum possible speed, what can you do ?

Answer)We need to use –direct argument in import command to use direct import fast path and this –direct can be used only with MySQL and PostGreSQL as of now.

25)While connecting to MySQL through Sqoop, I am getting Connection Failure exception what might be the root cause and fix for this error scenario?

Answer)This might be due to insufficient permissions to access your MySQL database over the network. To confirm this we can try the below command to connect to MySQL database from Sqoop's client machine.
$ mysql --host=MySql node > --database=test --user= --password=
If this is the case then we need grant permissions user @ sqoop client machine as per the answer to Question 6 in this post.

26)What is the importance of eval tool?

Answer)It allow users to run sample SQL queries against Database and preview the result on the console.

27)What is the process to perform an incremental data load in Sqoop?

Answer)The process to perform incremental data load in Sqoop is to synchronize the modified or updated data (often referred as delta data) from RDBMS to Hadoop. The delta data can be facilitated through the incremental load command in Sqoop.
Incremental load can be performed by using Sqoop import command or by loading the data into hive without overwriting it. The different attributes that need to be specified during incremental load in Sqoop are-
1)Mode (incremental) –The mode defines how Sqoop will determine what the new rows are. The mode can have value as Append or Last Modified.
2)Col (Check-column) –This attribute specifies the column that should be examined to find out the rows to be imported.
3)Value (last-value) –This denotes the maximum value of the check column from the previous import operation.

28)What is the significance of using –compress-codec parameter?

www.smartdatacamp.com

Answer)To get the out file of a sqoop import in formats other than .gz like .bz2 we use the –compress -code parameter.

29)Can free form SQL queries be used with Sqoop import command? If yes, then how can they be used?

Answer)Sqoop allows us to use free form SQL queries with the import command. The import command should be used with the –e and – query options to execute free form SQL queries. When using the –e and –query options with the import command the –target dir value must be specified.

30)What is the purpose of sqoop-merge?

Answer)The merge tool combines two datasets where entries in one dataset should overwrite entries of an older dataset preserving only the newest version of the records between both the data sets.

31)How do you clear the data in a staging table before loading it by Sqoop?

Answer)By specifying the –clear-staging-table option we can clear the staging table before it is loaded. This can be done again and again till we get proper data in staging.

32)How will you update the rows that are already exported?

Answer)The parameter –update-key can be used to update existing rows. In it a comma-separated list of columns is used which uniquely identifies a row. All of these columns is used in the WHERE clause of the generated UPDATE query. All other table columns will be used in the SET part of the query.

33)What is the role of JDBC driver in a Sqoop set up?

Answer)To connect to different relational databases sqoop needs a connector. Almost every DB vendor makes this connecter available as a JDBC driver which is specific to that DB. So Sqoop needs the JDBC driver of each of the database it needs to interact with.

34)When to use --target-dir and when to use --warehouse-dir while importing data?

Answer)To specify a particular directory in HDFS use --target-dir but to specify the parent directory of all the sqoop jobs use --warehouse-dir. In this case under the parent directory sqoop will cerate a directory with the same name as th e table.

35)When the source data keeps getting updated frequently, what is the approach to keep it in sync with the data in HDFS imported by sqoop?

Answer)sqoop can have 2 approaches.
a − To use the --incremental parameter with append option where value of some columns are checked and only in case of modified values the row is imported as a new row.
b − To use the --incremental parameter with lastmodified option where a date column in the source is checked for records which have been updated after the last import.

36)Is it possible to add a parameter while running a saved job?

Answer)Yes, we can add an argument to a saved job at runtime by using the --exec option
sqoop job --exec jobname -- -- newparameter

37)Before starting the data transfer using mapreduce job, sqoop takes a long time to retrieve the minimum and maximum values of columns mentioned in –split-by parameter. How can we make it efficient?

Answer)We can use the --boundary –query parameter in which we specify the min and max value for the column based on which the split can happen into multiple mapreduce tasks. This makes it faster as the query inside the –boundary-query parameter is executed first and the job is ready with the information on how many mapreduce tasks to create before executing the main query.

38)How will you implement all-or-nothing load using sqoop?

Answer)Using the staging-table option we first load the data into a staging table and then load it to the final target table only if the staging load is successful.

39)How will you update the rows that are already exported?

Answer)The parameter --update-key can be used to update existing rows. In it a comma-separated list of columns is used which uniquely identifies a row. All of these columns is used in the WHERE clause of the generated UPDATE query. All other table columns will be used in the SET part of the query.

www.smartdatacamp.com

40)How can you sync a exported table with HDFS data in which some rows are deleted?

Answer)Truncate the target table and load it again.

41)How can we load to a column in a relational table which is not null but the incoming value from HDFS has a null value?

Answer)By using the –input-null-string parameter we can specify a default value and that will allow the row to be inserted into the target table.

42)How can you schedule a sqoop job using Oozie?

Answer)Oozie has in-built sqoop actions inside which we can mention the sqoop commands to be executed.

43)Sqoop imported a table successfully to HBase but it is found that the number of rows is fewer than expected. What can be the cause?

Answer)Some of the imported records might have null values in all the columns. As Hbase does not allow all null values in a row, those rows get dropped.

44)How can you force sqoop to execute a free form Sql query only once and import the rows serially.

Answer)By using the –m 1 clause in the import command, sqoop cerates only one mapreduce task which will import the rows sequentially.

45)In a sqoop import command you have mentioned to run 8 parallel Mapreduce task but sqoop runs only 4. What can be the reason?

Answer)The Mapreduce cluster is configured to run 4 parallel tasks. So the sqoop command must have number of parallel tasks less or equal to that of the MapReduce cluster.

46)What happens when a table is imported into a HDFS directory which already exists using the –apend parameter?

Answer)Using the --append argument, Sqoop will import data to a temporary directory and then rename the files into the normal target directory in a manner that does not conflict with existing filenames in that directory.

47)How to import only the updated rows form a table into HDFS using sqoop assuming the source has last update timestamp details for each row?

Answer)By using the lastmodified mode. Rows where the check column holds a timestamp more recent than the timestamp specified with --last-value are imported.

48)What does the following query do?
**$ sqoop import --connect jdbc:mysql://host/dbname --table EMPLOYEES **
--where start_date > 2012-11-09

Answer)It imports the employees who have joined after 9-Nov-2012.

49)Give a Sqoop command to import all the records from employee table divided into groups of records by the values in the column department_id.

Answer) $ sqoop import --connect jdbc:mysql://db.foo.com/corp --table EMPLOYEES \

--split-by dept_id

**50)What does the following query do? $ sqoop import --connect jdbc:mysql://db.foo.com/somedb --table sometable **
--where "id > 1000" --target-dir /incremental_dataset --append

Answer)It performs an incremental import of new data, after having already imported the first 1000 rows of a table

www.smartdatacamp.com

Apache Flume

Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms.

1)Explain about the core components of Flume?

Answer)The core components of Flume are –
Event-The single log entry or unit of data that is transported.
Source-This is the component through which data enters Flume workflows.
Sink-It is responsible for transporting data to the desired destination.
Channel-it is the duct between the Sink and Source.
Agent-Any JVM that runs Flume.
Client-The component that transmits event to the source that operates with the agent.

2)Can I run two instances of the flume node on the same unix machine?

Answer)Yes. Run flume with the -n option.
flume node
flume node -n physicalnodename

3)I'm generating events from my application and sending it to a flume agent listening for Thrift/Avro RPCs and my timestamps seem to be in the 1970s.

Answer)Event generated is expected to have unix time in milliseconds. If the data is being generated by an external application, this application must generated data in terms of milliseconds.
For example, 1305680461000 should result in 5/18/11 01:01:01 GMT, but 1305680461 will result in something like 1/16/70 2:41:20 GMT

4)Can I control the level of HDFS replication / block size / other client HDFS property?

Answer)Yes. HDFS block size and replication level are HDFS client parameters, so you should expect them to be set by client. The parameters you get are probably coming from hadoop-core.*.jar file (it usually contains hdfs-default.xml and friends). If you want to overwrite the default parameters, you need to set dfs.block.size and dfs.replication in your hdfs-site.xml or flume-site.xml file

5)Which is the reliable channel in Flume to ensure that there is no data loss?

Answer) FILE Channel is the most reliable channel among the 3 channels JDBC, FILE and MEMORY.

6)How multi-hop agent can be setup in Flume?

Answer)Avro RPC Bridge mechanism is used to setup Multi-hop agent in Apache Flume.

7)Does Apache Flume provide support for third party plug-ins?

Answer)Most of the data analysts use Apache Flume has plug-in based architecture as it can load data from external sources and transfer it to external destinations.

8)Is it possible to leverage real time analysis on the big data collected by Flume directly? If yes, then explain how.

Answer)Data from Flume can be extracted, transformed and loaded in real-time into Apache Solr servers using MorphlineSolrSink

9)What is a channel?

Answer)It stores events,events are delivered to the channel via sources operating within the agent.An event stays in the channel until a sink removes it for further transport.

10)What is Interceptor?

Answer)An interceptor can modify or even drop events based on any criteria chosen by the developer.

11)Explain about the replication and multiplexing selectors in Flume.

Answer)Channel Selectors are used to handle multiple channels. Based on the Flume header value, an event can be written just to a single channel or to multiple channels. If a channel selector is not specified to the source then by default it is the Replicating selector. Using the replicating selector, the same event is written to all the channels in the source's channels list.

www.smartdatacamp.com

Multiplexing channel selector is used when the application has to send different events to different channels.

12)Does Apache Flume provide support for third party plug-ins?

Answer)Most of the data analysts use Apache Flume has plug-in based architecture as it can load data from external sources and transfer it to external destinations.

13)Agent communicate with other Agents?

Answer)NO each agent runs independently. Flume can easily horizontally. As a result ther is no single point of failure.

14)what are the complicated steps in Flume configurations?

Answer)Flume can processing straming data. so if started once, there is no stop/end to the process. asynchronously it can flows data from source to HDFS via agent. First of all agent should know individual components how they are connecto to load data. so configuraton is trigger to load streaming data. for example consumerkey, consumersecret accessToken and accessTokenSecret are key factor to download data from twitter.

15)What is flume agent?

Answer)A flume agent is JVM holds the flume core components(source, channel, sink) through which events flow from an external source like web-servers to destination like HDFS. Agent is heart of the Apache Flime.

16)What is Flume event?

Answer)A unit of data with set of string attribute called Flume event. The external source like web-server send events to the source. Internally Flume has inbuilt functionality to understand the source format.Each log file is consider as an event. Each event has header and value sectors, which has header information and appropriate value that assign to articular header.

17)Is it possible to leverage real time analysis on the big data collected by Flume directly? If yes, then explain how.

Answer)Data from Flume can be extracted, transformed and loaded in real-time into Apache Solr servers using MorphlineSolrSink

18)Differentiate between FileSink and FileRollSink

Answer)The major difference between HDFS FileSink and FileRollSink is that HDFS File Sink writes the events into the Hadoop Distributed File System (HDFS) whereas File Roll Sink stores the events into the local file system?

19)How to use exec source?

Answer)Set the agents source type property to exec as below.
agents.sources.sourceid.type=exec

20)How to improve performance?

Answer)Batching the events: You can specify the number of events to be written per transaction by changing the batch size, which has default value as 20
Agent.sources.sourceid.batchSize=2000

21)Why to provide higher value in batchSize?

Answer)When your input data is large and you find that you can not write to your channel fast enough.Having bigger batch size will reduce the overall average transaction overhead per event. However, you must have tested it before deciding batch size.

22)What is the Problem with SpoolDir?

Answer)Whenever, due to error or any other reason Flume, restarts it will create duplicate events on any files in the spooling directory that are re-transmitted due to not being marked as finished.

23)How can I tell if I have a library loaded when flume runs?

Answer)From the command line, you can run flume classpath to see the jars and the order Flume is attempting to load them in.

24)How can I tell if a plugin has been loaded by a flume node?

www.smartdatacamp.com

Answer)You can look at the node's plugin status web page – http://<master>:35871/extension.jsp Alternately, you can look at the logs.

25)Why does the master need to have plugins installed?

Answer)The master needs to have plugins installed in order to validate configs it is sending to nodes.

26)How can I tell if a plugin has been loaded by a flume master?

Answer)You can look at the node's plugin status web page – http://<master>:35871/masterext.jsp Alternately, you can look at the logs.

27)How can I tell if a plugin has been loaded by a flume node?

Answer)You can look at the node's plugin status web page – http://<node>:35862/staticconfig.jsp Alternately, you can look at the logs.

28)How can I tell if my flume-site.xml configuration values are being read properly?

Answer)You can go to the node or master's static config web page to see what configuration values are loaded. http://<node>:35862/staticconfig.jsp http://<master>:35871/masterstaticconfig.jsp

29)I'm having a hard time getting the LZO codec to work.

Answer)Flume by default reads the $HADOOP_CONF_DIR/core-site.xml which may have the io.compression.codecs setting set. You can make the settting <final> so that flume does not attempt to override the setting.

29)I lose my configurations when I restart the master. What's happening?

Answer)The default path to write information is set to this value. You may want to override this to a place will be persistent across reboots such as /var/lib/flume. <property> <name>flume.master.zk.logdir</name> <value>/tmp/flume-$ Unknown macro: {user.name}

-zk</value> <description>The base directory in which the ZBCS stores data.</description> </property>

30)How can I get metrics from a node?

Answer)Flume nodes report metrics which you can use to debug and to see progress. You can look at a node's status web page by pointing your browser to port 35862. (http://<node>:35862).

31)How can I tell if data is arriving at the collector?

Answer)When events arrive at a collector, the source counters should be incremented on the node's metric page. For example, if you have a node called foo you should see the following fields have growing values when you refresh the page.
LogicalNodeManager.foo.source.CollectorSource.number of bytes
LogicalNodeManager.foo.source.CollectorSource.number of events

32)How can I tell if data is being written to HDFS?

Answer)Data in hdfs doesn't "arrive" in hdfs until the file is closed or certain size thresholds are met. As events are written to hdfs, the sink counters on the collector's metric page should be incrementing. In particular look for fields that match the following names: *.Collector.GunzipDecorator.UnbatchingDecorator.AckChecksumChecker.InsistentAppend.append* *.appendSuccesses are successful writes. If other values like appendRetries or appendGiveups are incremented, they indicate a problem with the attemps to write.

33)I am getting a lot of duplicated event data. Why is this happening and what can I do to make this go away?

Answer)tail/multiTail have been reported to restart file reads from the beginning of files if the modification rate reaches a certain rate. This is a fundamental problem with a non-native implementation of tail. A work around is to use the OS's tail mechanism in an exec source (exec("tail -n +0 -F filename")). Alternately many people have modified their applications to push to a Flume agent with an open rpc port such as syslogTcp or thriftSource, avroSource. In E2E mode, agents will attempt to retransmit data if no acks are recieved after flume.agent.logdir.retransmit milliseconds have expried (this is a flume-site.xml property). Acks do not return until after the collector's roll time, flume.collector.roll.millis , expires (this can be set in the flume-site.xml file or as an argument to a collector) . Make sure that the retry time on the agents is at least 2x that of the roll time on the collector. If that was in E2E mode goes down, it will attempt to recover and resend data that did not receive acknowledgements on restart. This may result in some duplicates.

www.smartdatacamp.com

34)I have encountered a "Could not increment version counter" error message.

Answer)This is a zookeeper issue that seems related to virtual machines or machines that change IP address while running. This should only occur in a development environment – the work around here is to restart the master.

35)I have encountered an IllegalArgumentException related to checkArgument and EventImpl.

Answer)Here's an example stack trace: 2011-07-11 01:12:34,773 ERROR com.cloudera.flume.core.connector.DirectDriver: Driving src/sink failed! LazyOpenSource | LazyOpenDecorator because null java.lang.IllegalArgumentException at com.google.common.base.Preconditions.checkArgument(Preconditions.java:75) at com.cloudera.flume.core.EventImpl.<init>(EventImpl.java:97) at com.cloudera.flume.core.EventImpl.<init>(EventImpl.java:87) at com.cloudera.flume.core.EventImpl.<init>(EventImpl.java:71) at com.cloudera.flume.handlers.syslog.SyslogWireExtractor.buildEvent(SyslogWireExtractor.java:120) at com.cloudera.flume.handlers.syslog.SyslogWireExtractor.extract(SyslogWireExtractor.java:192) at com.cloudera.flume.handlers.syslog.SyslogWireExtractor.extractEvent(SyslogWireExtractor.java:89) at com.cloudera.flume.handlers.syslog.SyslogUdpSource.next(SyslogUdpSource.java:88) at com.cloudera.flume.handlers.debug.LazyOpenSource.next(LazyOpenSource.java:57) at com.cloudera.flume.core.connector.DirectDriver$PumperThread.run(DirectDriver.java:89) This indicates an attempt to create an event body that is larger than the maximum allowed body size (default 32k). You can increase the size of the max event by setting flume.event.max.size.bytes in your flume-site.xml file to a larger value. We are addressing this with issue FLUME-712.

36)I'm getting OutOfMemoryExceptions in my collectors or agents.

Answer)Add -XX:+HeapDumpOnOutOfMemoryError to the JVM_MEM_OPTS env variable or flume-env.sh file. This should dump heap upon thse kinds of errors and allow you to determine what objects are consuming excessive memory by using the jhat java heap viewer program. There have been instances of queues that are unbounded. Several of these have been fixed in v0.9.5. There are situations where queue sizes are too large for certain messages. For example, if batching is used, each event can takes up more memory. The default queue size in thrift sources is 1000 items. With batching individual events can become megabytes in size which may cause memory exhaustion. For example making batches of 1000 1000-byte messages with a queue of 1000 events could result in flume requiring 1GB of memory! In these cases, reduced the size of the thrift queue to bound potential the memory usage by setting flume.thrift.queuesize

```
<property>
 <name>flume.thrift.queuesize</name>
 <value>500</value>
 </property>
```

Apache Cassandra

Apache Cassandra is a free and open-source distributed NoSQL database management system designed to handle large amounts of data across many commodity servers, providing high availability with no single point of failure.

1)Explain what is Cassandra?

Answer)Cassandra is an open source data storage system developed at Facebook for inbox search and designed for storing and managing large amounts of data across commodity servers. It can server as both
Real time data store system for online applications
Also as a read intensive database for business intelligence system

2)What do you understand by Commit log in Cassandra?

Answer)Commit log is a crash-recovery mechanism in Cassandra. Every write operation is written to the commit log.

3)In which language Cassandra is written?

Answer)Cassandra is written in Java. It is originally designed by Facebook consisting of flexible schemas. It is highly scalable for big data.

4)List the benefits of using Cassandra.

Answer)Unlike traditional or any other database, Apache Cassandradelivers near real-time performance simplifying the work of Developers, Administrators, Data Analysts and Software Engineers.
Instead of master-slave architecture, Cassandra is established on peer-to-peer architecture ensuring no failure.
It also assures phenomenal flexibility as it allows insertion of multiple nodes to any Cassandra cluster in any datacenter. Further, any client can forward its request to any server.
Cassandra facilitates extensible scalability and can be easily scaled up and scaled down as per the requirements. With a high throughput for read and write operations, this NoSQL application need not be restarted while scaling.
Cassandra is also revered for its strong data replication on nodes capability as it allows data storage at multiple locations enabling users to retrieve data from another location if one node fails. Users have the option to set up the number of replicas they want to create.

www.smartdatacamp.com

Shows brilliant performance when used for massive datasets and thus, the most preferable NoSQL DB by most organizations.
Operates on column-oriented structure and thus, quickens and simplifies the process of slicing. Even data access and retrieval becomes more efficient with column-based data model.
Further, Apache Cassandra supports schema-free/schema-optional data model, which un-necessitate the purpose of showing all the columns required by your application.Find out how Cassandra Versus MongoDB can help you get ahead in your career!

5)What was the design goal of Cassandra?

Answer)The main design goal of Cassandra was to handle big data workloads across multiple nodes without a single point of failure.

6)What are the main components of Cassandra data models?

Answer)Following are the main components of Cassandra data model:
Cluster
Keyspace
Column
Column and Family

7)What are the other components of Cassandra?

Answer)Some other components of Cassandra are:
Node
Data Center
Commit log
Mem-table
SSTable
Bloom Filter

8)Explain the concept of Tunable Consistency in Cassandra?

Answer)Tunable Consistency is a phenomenal characteristic that makes Cassandra a favored database choice of Developers, Analysts and Big data Architects. Consistency refers to the up-to-date and synchronized data rows on all their replicas. Cassandra's Tunable Consistency allows users to select the consistency level best suited for their use cases. It supports two consistencies -Eventual and Consistency and Strong Consistency.

The former guarantees consistency when no new updates are made on a given data item, all accesses return the last updated value eventually. Systems with eventual consistency are known to have achieved replica convergence.

For Strong consistency, Cassandra supports the following condition:
R + W > N, where
N – Number of replicas
W – Number of nodes that need to agree for a successful write
R – Number of nodes that need to agree for a successful read

9)How does Cassandra write?

Answer)Cassandra performs the write function by applying two commits-first it writes to a commit log on disk and then commits to an in-memory structured known as memtable. Once the two commits are successful, the write is achieved. Writes are written in the table structure as SSTable (sorted string table). Cassandra offers speedier write performance.

10)Why cant I set listen_address to listen on 0.0.0.0 (all my addresses)?

Answer)Cassandra is a gossip-based distributed system and listen_address is the address a node tells other nodes to reach it at. Telling other nodes "contact me on any of my addresses" is a bad idea; if different nodes in the cluster pick different addresses for you, Bad Things happen.

If you don't want to manually specify an IP to listen_address for each node in your cluster (understandable!), leave it blank and Cassandra will use InetAddress.getLocalHost() to pick an address. Then it's up to you or your ops team to make things resolve correctly (/etc/hosts/, dns, etc).

One exception to this process is JMX, which by default binds to 0.0.0.0

11)What ports does Cassandra use?

Answer)By default, Cassandra uses 7000 for cluster communication (7001 if SSL is enabled), 9042 for native protocol clients, and 7199 for JMX. The internode communication and native protocol ports are configurable in the Cassandra Configuration File. The JMX port is configurable in cassandra-env.sh (through JVM options). All ports are TCP.

12)Define Mem-table in Cassandra.

Answer) It is a memory-resident data structure. After commit log, the data will be written to the mem-table. Mem-table is in-memory/write-back cache space consisting of content in key and column format. The data in mem- table is sorted by key, and each column family consists of a distinct mem-table that retrieves column data via key. It stores the writes until it is full, and then flushed out.

www.smartdatacamp.com

13)What is SSTable?

Answer)SSTable or 'Sorted String Table,' refers to an important data file in Cassandra. It accepts regular written memtables which are stored on disk and exist for each Cassandra table. Being immutable, SStables do not allow any further addition and removal of data items once written. For each SSTable, Cassandra creates three separate files like partition index, partition summary and a bloom filter.

14)What is bloom filter?

Answer)Bloom filter is an off-heap data structure to check whether there is any data available in the SSTable before performing any I/O disk operation.

15)Establish the difference between a node, cluster and data centres in Cassandra.

Answer)Node is a single machine running Cassandra.
Cluster is a collection of nodes that have similar type of data grouped together.
Data centres are useful components when serving customers in different geographical areas. Different nodes of a cluster are grouped into different data centres.

16)Define composite type in Cassandra?

Answer)In Cassandra, composite type allows to define a key or a column name with a concatenation of data of different type. You can use two types of Composite Types:
Row Key
Column Name

17)What is keyspace in Cassandra?

Answer)In Cassandra, a keyspace is a namespace that determines data replication on nodes. A cluster contains of one keyspace per node.

18)Define the management tools in Cassandra.

Answer)DataStaxOpsCenter: internet-based management and monitoring solution for Cassandra cluster and DataStax. It is free to download and includes an additional Edition of OpsCenter SPM primarily administers Cassandra metrics and various OS and JVM metrics. Besides Cassandra, SPM also monitors Hadoop, Spark, Solr, Storm, zookeeper and other Big Data platforms. The main features of SPM include correlation of events and metrics, distributed

transaction tracing, creating real-time graphs with zooming, anomaly detection and heartbeat alerting.

19)Explain CAP Theorem.

Answer)With a strong requirement to scale systems when additional resources are needed, CAP Theorem plays a major role in maintaining the scaling strategy. It is an efficient way to handle scaling in distributed systems. Consistency Availability and Partition tolerance (CAP) theorem states that in distributed systems like Cassandra, users can enjoy only two out of these three characteristics.
One of them needs to be sacrificed. Consistency guarantees the return of most recent write for the client, Availability returns a rational response within minimum time and in Partition Tolerance, the system will continue its operations when network partitions occur. The two options available are AP and CP.

20)How to write a query in Cassandra?

Answer)Using CQL (Cassandra Query Language).Cqlsh is used for interacting with database.

21)What OS Cassandra supports?

Answer)Windows and Linux

22)Talk about the concept of tunable consistency in Cassandra?

Answer)Tunable Consistency is a characteristic that makes Cassandra a favored database choice of Developers, Analysts and Big data Architects. Consistency refers to the up-to-date and synchronized data rows on all their replicas. Cassandra's Tunable Consistency allows users to select the consistency level best suited for their use cases. It supports two consistencies – Eventual Consistency and Strong Consistency.

23)What are the three components of Cassandra write?

Answer)The three components are:
Commitlog write
Memtable write
SStable write
Cassandra first writes data to a commit log and then to an in-memory table structure memtable and at last in SStable.

www.smartdatacamp.com

24)What is the syntax to create keyspace in Cassandra?

Answer)CREATE KEYSPACE identifier WITH properties

25)What happens to existing data in my cluster when I add new nodes?

Answer)When a new nodes joins a cluster, it will automatically contact the other nodes in the cluster and copy the right data to itself.

26)When I delete data from Cassandra, but disk usage stays the same

Answer)Data you write to Cassandra gets persisted to SSTables. Since SSTables are immutable, the data can't actually be removed when you perform a delete, instead, a marker (also called a tombstone) is written to indicate the value's new status. Never fear though, on the first compaction that occurs between the data and the tombstone, the data will be expunged completely and the corresponding disk space recovered

27)Explain zero consistency.

Answer)In zero consistency the write operations will be handled in the background, asynchronously. It is the fastest way to write data.

28)What do you understand by Kundera?

Answer)Kundera is an object-relational mapping (ORM) implementation for Cassandra which is written using Java annotations.

29)Why does nodetool ring only show one entry, even though my nodes logged that they see each other joining the ring?

Answer)This happens when you have the same token assigned to each node. Don't do that. Most often this bites people who deploy by installing Cassandra on a VM (especially when using the Debian package, which auto-starts Cassandra after installation, thus generating and saving a token), then cloning that VM to other nodes. The easiest fix is to wipe the data and commitlog directories, thus making sure that each node will generate a random token on the next restart.

30)Can I change the replication factor (a a keyspace) on a live cluster?

Answer)Yes, but it will require running a full repair (or cleanup) to change the replica count of existing data:
Alter the replication factor for desired keyspace (using cqlsh for instance).
If you're reducing the replication factor, run nodetool cleanup on the cluster to remove surplus replicated data. Cleanup runs on a per-node basis.
If you're increasing the replication factor, run nodetool repair -full to ensure data is replicated according to the new configuration. Repair runs on a per-replica set basis. This is an intensive process that may result in adverse cluster performance. It's highly recommended to do rolling repairs, as an attempt to repair the entire cluster at once will most likely swamp it. Note that you will need to run a full repair (-full) to make sure that already repaired sstables are not skipped.

31)Can I Store (large) BLOBs in Cassandra?

Answer)Cassandra isnt optimized for large file or BLOB storage and a single blob value is always read and send to the client entirely. As such, storing small blobs (less than single digit MB) should not be a problem, but it is advised to manually split large blobs into smaller chunks.
Please note in particular that by default, any value greater than 16MB will be rejected by Cassandra due the max_mutation_size_in_kb configuration of the Cassandra Configuration File file (which default to half of commitlog_segment_size_in_mb, which itself default to 32MB).

32)Nodetool says "Connection refused to host: 127.0.1.1" for any remote host.How to fix it?

Answer)Nodetool relies on JMX, which in turn relies on RMI, which in turn sets up its own listeners and connectors as needed on each end of the exchange. Normally all of this happens behind the scenes transparently, but incorrect name resolution for either the host connecting, or the one being connected to, can result in crossed wires and confusing exceptions.
If you are not using DNS, then make sure that your /etc/hosts files are accurate on both ends. If that fails, try setting the -Djava.rmi.server.hostname=public name JVM option near the bottom of cassandra-env.sh to an interface that you can reach from the remote machine.

33)Will batching my operations speed up my bulk load?

Answer)No. Using batches to load data will generally just add spikes of latency. Use asynchronous INSERTs instead, or use true Bulk Loading.
An exception is batching updates to a single partition, which can be a Good Thing (as long as the size of a single batch stay reasonable). But never ever blindly batch everything.

34)Why does top report that Cassandra is using a lot more memory than the Java heap max?

www.smartdatacamp.com

Answer)Cassandra uses Memory Mapped Files (mmap) internally. That is, we use the operating system's virtual memory system to map a number of on-disk files into the Cassandra process' address space. This will "use" virtual memory; i.e. address space, and will be reported by tools like top accordingly, but on 64 bit systems virtual address space is effectively unlimited so you should not worry about that.
What matters from the perspective of "memory use" in the sense as it is normally meant, is the amount of data allocated on brk() or mmap'd /dev/zero, which represent real memory used. The key issue is that for a mmap'd file, there is never a need to retain the data resident in physical memory. Thus, whatever you do keep resident in physical memory is essentially just there as a cache, in the same way as normal I/O will cause the kernel page cache to retain data that you read/write.
The difference between normal I/O and mmap() is that in the mmap() case the memory is actually mapped to the process, thus affecting the virtual size as reported by top. The main argument for using mmap() instead of standard I/O is the fact that reading entails just touching memory - in the case of the memory being resident, you just read it - you don't even take a page fault (so no overhead in entering the kernel and doing a semi-context switch).

35)What are seeds?

Answer)Seeds are used during startup to discover the cluster.
If you configure your nodes to refer some node as seed, nodes in your ring tend to send Gossip message to seeds more often (also see the section on gossip) than to non-seeds. In other words, seeds are worked as hubs of Gossip network. With seeds, each node can detect status changes of other nodes quickly.
Seeds are also referred by new nodes on bootstrap to learn other nodes in ring. When you add a new node to ring, you need to specify at least one live seed to contact. Once a node join the ring, it learns about the other nodes, so it doesn't need seed on subsequent boot.
You can make a seed a node at any time. There is nothing special about seed nodes. If you list the node in seed list it is a seed
Seeds do not auto bootstrap (i.e. if a node has itself in its seed list it will not automatically transfer data to itself) If you want a node to do that, bootstrap it first and then add it to seeds later. If you have no data (new install) you do not have to worry about bootstrap at all.
Recommended usage of seeds:
pick two (or more) nodes per data center as seed nodes.
sync the seed list to all your nodes

36)Does single seed mean single point of failure?

Answer)The ring can operate or boot without a seed; however, you will not be able to add new nodes to the cluster. It is recommended to configure multiple seeds in production system.

37)Why do I see messages dropped in the logs?

Answer)This is a symptom of load shedding Cassandra defending itself against more requests than it can handle.
Internode messages which are received by a node, but do not get not to be processed within their proper timeout (see read_request_timeout, write_request_timeout, in the Cassandra Configuration File), are dropped rather than processed (since the as the coordinator node will no longer be waiting for a response).
For writes, this means that the mutation was not applied to all replicas it was sent to. The inconsistency will be repaired by read repair, hints or a manual repair. The write operation may also have timeouted as a result.
For reads, this means a read request may not have completed.
Load shedding is part of the Cassandra architecture, if this is a persistent issue it is generally a sign of an overloaded node or cluster.

38)Cassandra dies with java.lang.OutOfMemoryError: Map failed

Answer)If Cassandra is dying specifically with the "Map failed" message, it means the OS is denying java the ability to lock more memory. In linux, this typically means memlock is limited. Check /proc/pid of cassandra/limits to verify this and raise it (eg, via ulimit in bash). You may also need to increase vm.max_map_count. Note that the debian package handles this for you automatically.

39)What happens if two updates are made with the same timestamp?

Answer)Updates must be commutative, since they may arrive in different orders on different replicas. As long as Cassandra has a deterministic way to pick the winner (in a timestamp tie), the one selected is as valid as any other, and the specifics should be treated as an implementation detail. That said, in the case of a timestamp tie, Cassandra follows two rules: first, deletes take precedence over inserts/updates. Second, if there are two updates, the one with the lexically larger value is selected.

40)Why bootstrapping a new node fails with a "Stream failed" error?

Answer)Two main possibilities:
the GC may be creating long pauses disrupting the streaming process
compactions happening in the background hold streaming long enough that the TCP connection fails
In the first case, regular GC tuning advices apply. In the second case, you need to set TCP keepalive to a lower value (default is very high on Linux). Try to just run the following:
$ sudo /sbin/sysctl -w net.ipv4.tcp_keepalive_time=60 net.ipv4.tcp_keepalive_intvl=60 net.ipv4.tcp_keepalive_probes=5
To make those settings permanent, add them to your /etc/sysctl.conf file.

www.smartdatacamp.com

41)What is the concept of SuperColumn in Cassandra?

Answer)Cassandra SuperColumn is a unique element consisting of similar collections of data. They are actually key-value pairs with values as columns. It is a sorted array of columns, and they follow a hierarchy when in action.

42)When do you have to avoid secondary indexes?

Answer)Try not using secondary indexes on columns containing a high count of unique values as that will produce few results.

43)Mention what does the shell commands Capture and Consistency determines?

Answer)There are various Cqlsh shell commands in Cassandra. Command Capture, captures the output of a command and adds it to a file while, command Consistency display the current consistency level or set a new consistency level.

44)What is mandatory while creating a table in Cassandra?

Answer)While creating a table primary key is mandatory, it is made up of one or more columns of a table.

45)Mention what is Cassandra- CQL collections?

Answer)Cassandra CQL collections help you to store multiple values in a single variable. In Cassandra, you can use CQL collections in following ways
List: It is used when the order of the data needs to be maintained, and a value is to be stored multiple times (holds the list of unique elements)
SET: It is used for group of elements to store and returned in sorted orders (holds repeating elements)
MAP: It is a data type used to store a key-value pair of elements

46)Explain how Cassandra delete Data?

Answer)SSTables are immutable and cannot remove a row from SSTables. When a row needs to be deleted, Cassandra assigns the column value with a special value called Tombstone. When the data is read, the Tombstone value is considered as deleted.

47)Does Cassandra support ACID transactions?

Answer)Unlike relational databases, Cassandra does not support ACID transactions.

48)List the steps in which Cassandra writes changed data into commitlog?

Answer)Cassandra concatenates changed data to commitlog. Then Commitlog acts as a crash recovery log for data. Until the changed data is concatenated to commitlog, write operation will never be considered successful.

49)What is the use of ResultSet execute(Statement statement) method?

Answer)This method is used to execute a query. It requires a statement object.

50)What is Thrift?

Answer)Thrift is the name of the Remote Procedure Call (RPC) client used to communicate with the Cassandra server.

51)What is the use of "void close()" method?

Answer)This method is used to close the current session instance.

52)What are the main features of SPM in Cassandra?

Answer)The main features of SPM are
Correlation of events and metrics
Distributed transaction tracing
Creating real-time graphs with zooming
Detection and heartbeat alerting

53)When can you use ALTER KEYSPACE?

Answer)The ALTER KEYSPACE can be used to change properties such as the number of replicas and the durable_write of a keyspace.

www.smartdatacamp.com

54)What is Hector in Cassandra?

Answer)Hector was one of the early Cassandra clients. It is an open source project written in Java using the MIT license.

55)What do you understand by Snitches?

Answer)A snitch determines which data centers and racks nodes belong to. They inform Cassandra about the network topology so that requests are routed efficiently and allows Cassandra to distribute replicas by grouping machines into data centers and racks. Specifically, the replication strategy places the replicas based on the information provided by the new snitch. All nodes must return to the same rack and data center. Cassandra does its best not to have more than one replica on the same rack.

www.smartdatacamp.com

Apache HBase

Apache HBase is an open-source, non-relational, distributed database modeled after Google's Bigtable and is written in Java. It is developed as part of Apache Software Foundation's Apache Hadoop project and runs on top of HDFS (Hadoop Distributed File System), providing Bigtable-like capabilities for Hadoop.

1)When Should I Use HBase?

Answer)HBase isn't suitable for every problem.

First, make sure you have enough data. If you have hundreds of millions or billions of rows, then HBase is a good candidate. If you only have a few thousand/million rows, then using a traditional RDBMS might be a better choice due to the fact that all of your data might wind up on a single node (or two) and the rest of the cluster may be sitting idle.

Second, make sure you can live without all the extra features that an RDBMS provides (e.g., typed columns, secondary indexes, transactions, advanced query languages, etc.) An application built against an RDBMS cannot be "ported" to HBase by simply changing a JDBC driver, for example. Consider moving from an RDBMS to HBase as a complete redesign as opposed to a port.

Third, make sure you have enough hardware. Even HDFS doesn't do well with anything less than 5 DataNodes (due to things such as HDFS block replication which has a default of 3), plus a NameNode.

HBase can run quite well stand-alone on a laptop - but this should be considered a development configuration only.

2)Can you create HBase table without assigning column family?

Answer)No, Column family also impact how the data should be stored physically in the HDFS file system, hence there is a mandate that you should always have at least one column family. We can also alter the column families once the table is created.

3)Does HBase support SQL?

Answer)Not really. SQL-ish support for HBase via Hive is in development, however Hive is based on MapReduce which is not generally suitable for low-latency requests.

4)Why are the cells above 10MB not recommended for HBase?

www.smartdatacamp.com

Answer)Large cells don't fit well into HBase's approach to buffering data. First, the large cells bypass the MemStoreLAB when they are written. Then, they cannot be cached in the L2 block cache during read operations. Instead, HBase has to allocate on-heap memory for them each time. This can have a significant impact on the garbage collector within the RegionServer process.

5)How should I design my schema in HBase?

Answer)A good introduction on the strength and weaknesses modelling on the various non-rdbms datastores is to be found in Ian Varley's Master thesis, No Relation: The Mixed Blessings of Non-Relational Databases. It is a little dated now but a good background read if you have a moment on how HBase schema modeling differs from how it is done in an RDBMS. Also, read keyvalue for how HBase stores data internally, and the section on schema.casestudies.

The documentation on the Cloud Bigtable website, Designing Your Schema, is pertinent and nicely done and lessons learned there equally apply here in HBase land; just divide any quoted values by ~10 to get what works for HBase: e.g. where it says individual values can be ~10MBs in size, HBase can do similar perhaps best to go smaller if you can and where it says a maximum of 100 column families in Cloud Bigtable, think ~10 when modeling on HBase.

6)Can you please provide an example of "good de-normalization" in HBase and how its held consistent (in your friends example in a relational db, there would be a cascadingDelete)? As I think of the users table: if I delete an user with the userid='123', do I have to walk through all of the other users column-family "friends" to guaranty consistency?! Is de-normalization in HBase only used to avoid joins? Our webapp doesn't use joins at the moment anyway.

Answer)You lose any concept of foreign keys. You have a primary key, that's it. No secondary keys/indexes, no foreign keys.

It's the responsibility of your application to handle something like deleting a friend and cascading to the friendships. Again, typical small web apps are far simpler to write using SQL, you become responsible for some of the things that were once handled for you.

Another example of "good denormalization" would be something like storing a users "favorite pages". If we want to query this data in two ways: for a given user, all of his favorites. Or, for a given favorite, all of the users who have it as a favorite. Relational database would probably have tables for users, favorites, and userfavorites. Each link would be stored in one row in the userfavorites table. We would have indexes on both 'userid' and 'favoriteid' and could thus query it in both ways described above. In HBase we'd probably put a column in both the users table and the favorites table, there would be no link table.

That would be a very efficient query in both architectures, with relational performing better much better with small datasets but less so with a large dataset.

Now asking for the favorites of these 10 users. That starts to get tricky in HBase and will undoubtedly suffer worse from random reading. The flexibility of SQL allows us to just ask the database for the answer to that question. In a small dataset it will come up with a decent

solution, and return the results to you in a matter of milliseconds. Now let's make that userfavorites table a few billion rows, and the number of users you're asking for a couple thousand. The query planner will come up with something but things will fall down and it will end up taking forever. The worst problem will be in the index bloat. Insertions to this link table will start to take a very long time. HBase will perform virtually the same as it did on the small table, if not better because of superior region distribution.

7)How would you design an Hbase table for many-to-many association between two entities, for example Student and Course?
I would define two tables:
Student: student id student data (name, address, ...) courses (use course ids as column qualifiers here) Course: course id course data (name, syllabus, ...) students (use student ids as column qualifiers here)
Does it make sense?

Answer)Your design does make sense.

As you said, you'd probably have two column-families in each of the Student and Course tables. One for the data, another with a column per student or course. For example, a student row might look like: Student : id/row/key = 1001 data:name = Student Name data:address = 123 ABC St courses:2001 = (If you need more information about this association, for example, if they are on the waiting list) courses:2002 = .

This schema gives you fast access to the queries, show all classes for a student (student table, courses family), or all students for a class (courses table, students family).

8)What is the maximum recommended cell size?

Answer)A rough rule of thumb, with little empirical validation, is to keep the data in HDFS and store pointers to the data in HBase if you expect the cell size to be consistently above 10 MB. If you do expect large cell values and you still plan to use HBase for the storage of cell contents, you'll want to increase the block size and the maximum region size for the table to keep the index size reasonable and the split frequency acceptable.

9)Why can't I iterate through the rows of a table in reverse order?

Answer)Because of the way HFile works: for efficiency, column values are put on disk with the length of the value written first and then the bytes of the actual value written second. To navigate through these values in reverse order, these length values would need to be stored twice (at the end as well) or in a side file. A robust secondary index implementation is the likely solution here to ensure the primary use case remains fast.

10)Can I fix OutOfMemoryExceptions in hbase?

www.smartdatacamp.com

Answer)Out-of-the-box, hbase uses a default of 1G heap size. Set the HBASE_HEAPSIZE environment variable in ${HBASE_HOME}/conf/hbase-env.sh if your install needs to run with a larger heap. HBASE_HEAPSIZE is like HADOOP_HEAPSIZE in that its value is the desired heap size in MB. The surrounding '-Xmx' and 'm' needed to make up the maximum heap size java option are added by the hbase start script (See how HBASE_HEAPSIZE is used in the ${HBASE_HOME}/bin/hbase script for clarification).

11)How do I enable hbase DEBUG-level logging?

Answer)Either add the following line to your log4j.properties file log4j.logger.org.apache.hadoop.hbase=DEBUG and restart your cluster or, if running a post-0.15.x version, you can set DEBUG via the UI by clicking on the 'Log Level' link (but you need set 'org.apache.hadoop.hbase' to DEBUG without the 'log4j.logger' prefix).

12)What ports does HBase use?

Answer)Not counting the ports used by hadoop - hdfs and mapreduce - by default, hbase runs the master and its informational http server at 60000 and 60010 respectively and regionservers at 60020 and their informational http server at 60030. ${HBASE_HOME}/conf/hbase-default.xml lists the default values of all ports used. Also check ${HBASE_HOME}/conf/hbase-site.xml for site-specific overrides.

13)Why is HBase ignoring HDFS client configuration such as dfs.replication?

Answer)If you have made HDFS client configuration on your hadoop cluster, HBase will not see this configuration unless you do one of the following:
Add a pointer to your HADOOP_CONF_DIR to CLASSPATH in hbase-env.sh or symlink your hadoop-site.xml from the hbase conf directory.
Add a copy of hadoop-site.xml to ${HBASE_HOME}/conf, or
If only a small set of HDFS client configurations, add them to hbase-site.xml
The first option is the better of the three since it avoids duplication.

14)Can I safely move the master from node A to node B?

Answer)Yes. HBase must be shutdown. Edit your hbase-site.xml configuration across the cluster setting hbase.master to point at the new location.

15)Can I safely move the hbase rootdir in hdfs?

Answer)Yes. HBase must be down for the move. After the move, update the hbase-site.xml across the cluster and restart.

16)How do I add/remove a node?

Answer)For removing nodes, see the section on decommissioning nodes in the HBase
Adding and removing nodes works the same way in HBase and Hadoop. To add a new node, do the following steps:
Edit $HBASE_HOME/conf/regionservers on the Master node and add the new address.
Setup the new node with needed software, permissions.
On that node run $HBASE_HOME/bin/hbase-daemon.sh start regionserver
Confirm it worked by looking at the Master's web UI or in that region server's log.
Removing a node is as easy, first issue "stop" instead of start then remove the address from the regionservers file.
For Hadoop, use the same kind of script (starts with hadoop-*), their process names (datanode, tasktracker), and edit the slaves file. Removing datanodes is tricky, please review the dfsadmin command before doing it.

17)Why do servers have start codes?

Answer)If a region server crashes and recovers, it cannot be given work until its lease times out. If the lease is identified only by an IP address and port number, then that server can't do any progress until the lease times out. A start code is added so that the restarted server can begin doing work immediately upon recovery

18)How do I monitor my HBase Cluster?

Answer)HBase emits performance metrics that you can monitor with Ganglia. Alternatively, you could use SPM for HBase

19)In which file the default configuration of HBase is stored.

Answer)hbase-site.xml

20)What is the RowKey.

www.smartdatacamp.com

Answer)Every row in an HBase table has a unique identifier called its rowkey (Which is equivalent to Primary key in RDBMS, which would be distinct throughout the table). Every interaction you are going to do in database will start with the RowKey only

21)Please specify the command (Java API Class) which you will be using to interact with HBase table.

Answer)Get, Put, Delete, Scan, and Increment

22)Which data type is used to store the data in HBase table column.

Answer)Byte Array,
Put p = new Put(Bytes.toBytes("John Smith"));
All the data in the HBase is stored as raw byte Array (10101010). Now the put instance is created which can be inserted in the HBase users table. © HadoopExam Leaning Resource

23)To locate the HBase data cell which three co-ordinate is used ?

Answer)HBase uses the coordinates to locate a piece of data within a table. The RowKey is the first coordinate. Following three co-ordinates define the location of the cell.
1.RowKey
2.Column Family (Group of columns)
3.Column Qualifier (Name of the columns or column itself e.g. Name, Email, Address) © HadoopExam Leaning Resource
Co-ordinates for the John Smith Name Cell.
["John Smith userID", “info”, “name”]

24)When you persist the data in HBase Row, In which tow places HBase writes the data to make sure the durability.

Answer)HBase receives the command and persists the change, or throws an exception if the write fails.
When a write is made, by default, it goes into two places:
a. the write-ahead log (WAL), also referred to as the HLog
b. and the MemStore
The default behavior of HBase recording the write in both places is in order to maintain data durability. Only after the change is written to and confirmed in both places is the write considered complete.

25)What is MemStore?

Answer)The MemStore is a write buffer where HBase accumulates data in memory before a permanent write.
Its contents are flushed to disk to form an HFile when the MemStore fills up.
It doesn't write to an existing HFile but instead forms a new file on every flush.
There is one MemStore per column family. (The size of the MemStore is defined by the system-wide property in
hbase-site.xml called hbase.hregion.memstore.flush.size)

26)What is HFile ?

Answer)The HFile is the underlying storage format for HBase.
HFiles belong to a column family and a column family can have multiple HFiles.
But a single HFile can't have data for multiple column families. © HadoopExam.com Leaning Resource

27)How HBase Handles the write failure?

Answer)Failures are common in large distributed systems, and HBase is no exception.
Imagine that the server hosting a MemStore that has not yet been flushed crashes. You'll lose the data that was in memory but not yet persisted. HBase safeguards against that by writing to the WAL before the write completes. Every server that's part of the.
HBase cluster keeps a WAL to record changes as they happen. The WAL is a file on the underlying file system. A write isn't considered successful until the new WAL entry is successfully written.
This guarantee makes HBase as durable as the file system backing it. Most of the time, HBase is backed by the Hadoop Distributed Filesystem (HDFS). If HBase goes down, the data that was not yet flushed from the MemStore to the HFile can be recovered by replaying the WAL

28)Which of the API command you will use to read data from HBase.

Answer)Get
Get g = new Get(Bytes.toBytes("John Smith"));
Result r = usersTable.get(g);

29)What is the BlockCache?

Answer)HBase also use the cache where it keeps the most used data in JVM Heap, along side Memstore.The BlockCache is designed to keep frequently accessed data from the HFiles in memory so as to avoid disk reads. Each column family has its own BlockCache

The Block in BlockCache is the unit of data that HBase reads from disk in a single pass. The HFile is physically laid out as a sequence of blocks plus an index over those blocks.This means reading

www.smartdatacamp.com

a block from HBase requires only looking up that blocks location in the index and retrieving it from disk.

The block is the smallest indexed unit of data and is the smallest unit of data that can be read from disk.

30)BlockSize is configured on which level?

Answer)The block size is configured per column family, and the default value is 64 KB. You may want to tweak this value larger or smaller depending on your use case.

31)If your requirement is to read the data randomly from HBase User table. Then what would be your preference to keep blcok size.

Answer)Having smaller blocks creates a larger index and thereby consumes more memory. If you frequently perform sequential scans, reading many blocks at a time, you can afford a larger block size. This allows you to save on memory because larger blocks mean fewer index entries and thus a smaller index.

32)What is a block, in a BlockCache ?

Answer)The Block in BlockCache is the unit of data that HBase reads from disk in a single pass. The HFile is physically laid out as a sequence of blocks plus an index over those blocks.

This means reading a block from HBase requires only looking up that blocks location in the index and retrieving it from disk. The block is the smallest indexed unit of data and is the smallest unit of data that can be read from disk.

The block size is configured per column family, and the default value is 64 KB. You may want to tweak this value larger or smaller depending on your use case.

33)While reading the data from HBase, from which three places data will be reconciled before returning the value ?

Answer)a. Reading a row from HBase requires first checking the MemStore for any pending modifications.
b. Then the BlockCache is examined to see if the block containing this row has been recently accessed.
c. Finally, the relevant HFiles on disk are accessed.
d. Note that HFiles contain a snapshot of the MemStore at the point when it was flushed. Data for a complete row can be stored across multiple HFiles.
e. In order to read a complete row, HBase must read across all HFiles that might contain information for that row in order to compose the complete record.

34)Once you delete the data in HBase, when exactly they are physically removed?

Answer)During Major compaction, Because HFiles are immutable, it's not until a major compaction runs that these tombstone records are reconciled and space is truly recovered from deleted records.

35)Please describe minor compaction

Answer)Minor : A minor compaction folds HFiles together, creating a larger HFile from multiple smaller HFiles.

36)Please describe major compactation?

Answer)When a compaction operates over all HFiles in a column family in a given region, it's called a major compaction. Upon completion of a major compaction, all HFiles in the column family are merged into a single file

37)What is tombstone record?

Answer)The Delete command doesn't delete the value immediately. Instead, it marks the record for deletion. That is, a new tombstone record is written for that value, marking it as deleted. The tombstone is used to indicate that the deleted value should no longer be included in Get or Scan results.

38)Can major compaction manually triggered?

Answer)Major compactions can also be triggered (or a particular region) manually from the shell. This is a relatively expensive operation and isn't done often. Minor compactions, on the other hand, are relatively lightweight and happen more frequently.

39)Which process or component is responsible for managing HBase RegionServer?

Answer)HMaster is the implementation of the Master Server.The Master server is responsible for monitoring all RegionServer instances in the cluster, and is the interface for all metadata changes. In a distributed cluster, the Master typically runs on the NameNode.

www.smartdatacamp.com

40)Which component is responsible for managing and monitoring of Regions?

Answer)HRegionServer is the RegionServer implementation. It is responsible for serving and managing regions. In a distributed cluster, a RegionServer runs on a DataNode.

41)What is the use of HColumnDescriptor?

Answer)An HColumnDescriptor contains information about a column family such as the number of versions, compression settings, etc. It is used as input when creating a table or adding a column. It is used as input when creating a table or adding a column. Once set, the parameters that specify a column cannot be changed without deleting the column and recreating it. If there is data stored in the column, it will be deleted when the column is deleted.

42)What is Field swap/promotion?

Answer)You can move the timestamp field of the row key or prefix it with another field. This approach uses the composite row key concept to move the sequential, monotonously increasing timestamp to a secondary position in the row key. If you already have a row key with more than one field, you can swap them. If you have only the timestamp as the current row key, you need to promote another field from the column keys, or even the value, into the row key. There is also a drawback to moving the time to the right-hand side in the composite key: you can only access data, especially time ranges, for a given swapped or promoted field.

43)Please tell us Operational command in Hbase, we you have used?

Answer)There are five main command in HBase.
1. Get
2. Put
3. Delete
4. Scan
5. Increment

44)Write down the Java Code snippet to open a connection in Hbase?

Answer)If you are going to open connection with the help of Java API.
The following code provide the connection
Configuration myConf = HBaseConfiguration.create();
HTableInterface usersTable = new HTable(myConf, "users");

45)Explain what is the row key?

Answer)Row key is defined by the application. As the combined key is pre-fixed by the rowkey, it enables the application to define the desired sort order. It also allows logical grouping of cells and make sure that all cells with the same rowkey are co-located on the same server.

46)What is the Deferred Log Flush in HBase?

Answer)The default behavior for Puts using the Write Ahead Log (WAL) is that HLog edits will be written immediately. If deferred log flush is used, WAL edits are kept in memory until the flush period. The benefit is aggregated and asynchronous HLog- writes, but the potential downside is that if the RegionServer goes down the yet-to-be-flushed edits are lost. This is safer, however, than not using WAL at all with Puts.

Deferred log flush can be configured on tables via HTableDescriptor. The default value of hbase.regionserver.optionallogflushinterval is 1000ms.

47)Can you describe the HBase Client: AutoFlush ?

Answer)When performing a lot of Puts, make sure that setAutoFlush is set to false on your HTable instance. Otherwise, the Puts will be sent one at a time to the RegionServer. Puts added via htable.add(Put) and htable.add(List Put) wind up in the same write buffer. If autoFlush = false, these messages are not sent until the write-buffer is filled. To explicitly flush the messages, call flushCommits. Calling close on the HTable instance will invoke flushCommits.

www.smartdatacamp.com

Apache ZooKeeper

Apache ZooKeeper is a software project of the Apache Software Foundation. It is essentially a distributed hierarchical key-value store, which is used to provide a distributed configuration service, synchronization service, and naming registry for large distributed systems.

1)What Is Zookeper??

Answer)ZooKeeper is a distributed co-ordination service to manage large set of hosts. Co-ordinating and managing a service in a distributed environment is a complicated process. ZooKeeper solves this issue with its simple architecture and API. ZooKeeper allows developers to focus on core application logic without worrying about the distributed nature of the application.

The ZooKeeper framework was originally built at "Yahoo!" for accessing their applications in an easy and robust manner. Later, Apache ZooKeeper became a standard for organized service used by Hadoop, HBase, and other distributed frameworks

2)What Are The Benefits Of Distributed Applications?

Answer)Reliability:Failure of a single or a few systems does not make the whole system to fail.
Scalability : Performance can be increased as and when needed by adding more machines with minor change in the configuration of the application with no downtime.
Transparency: Hides the complexity of the system and shows itself as a single entity / application.

3)What Are The Benefits Of Zookeeper?

Answer)Here are the benefits of using ZooKeeper:
Simple distributed coordination process
Synchronization:Mutual exclusion and co-operation between server processes. This process helps in Apache HBase for configuration management.
Ordered Messages
Serialization :Encode the data according to specific rules. Ensure your application runs consistently. This approach can be used in MapReduce to coordinate queue to execute running threads.
Reliability
Atomicity:Data transfer either succeed or fail completely, but no transaction is partial.

4)Explain The Types Of Znodes?

Answer)Znodes are categorized as persistence, sequential, and ephemeral.

Persistence znode - Persistence znode is alive even after the client, which created that particular znode, is disconnected. By default, all znodes are persistent unless otherwise specified.

Ephemeral znode - Ephemeral znodes are active until the client is alive. When a client gets disconnected from the ZooKeeper ensemble, then the ephemeral znodes get deleted automatically. For this reason, only ephemeral znodes are not allowed to have a children further. If an ephemeral znode is deleted, then the next suitable node will fill its position. Ephemeral znodes play an important role in Leader election.

Sequential znode - Sequential znodes can be either persistent or ephemeral. When a new znode is created as a sequential znode, then ZooKeeper sets the path of the znode by attaching a 10 digit sequence number to the original name. For example, if a znode with path /myapp is created as a sequential znode, ZooKeeper will change the path to /myapp0000000001 and set the next sequence number as 0000000002. If two sequential znodes are created concurrently, then ZooKeeper never uses the same number for each znode. Sequential znodes play an important role in Locking and Synchronization.

5)Explain The Zookeeper Workflow?

Answer)Once a ZooKeeper ensemble starts, it will wait for the clients to connect. Clients will connect to one of the nodes in the ZooKeeper ensemble. It may be a leader or a follower node. Once a client is connected, the node assigns a session ID to the particular client and sends an acknowledgement to the client. If the client does not get an acknowledgment, it simply tries to connect another node in the ZooKeeper ensemble. Once connected to a node, the client will send heartbeats to the node in a regular interval to make sure that the connection is not lost.

If a client wants to read a particular znode, it sends a read request to the node with the znode path and the node returns the requested znode by getting it from its own database. For this reason, reads are fast in ZooKeeper ensemble.

If a client wants to store data in the ZooKeeper ensemble, it sends the znode path and the data to the server. The connected server will forward the request to the leader and then the leader will reissue the writing request to all the followers. If only a majority of the nodes respond successfully, then the write request will succeed and a successful return code will be sent to the client. Otherwise, the write request will fail. The strict majority of nodes is called as Quorum.

6)Explain The Cli In Zookeeper?

Answer)ZooKeeper Command Line Interface (CLI) is used to interact with the ZooKeeper ensemble for development purpose. It is useful for debugging and working around with different options. To perform ZooKeeper CLI operations, first turn on your ZooKeeper server (“bin/zkServer.sh start”) and then, ZooKeeper client (bin/zkCli.sh).

Once the client starts, you can perform the following operation:
Create znodes

www.smartdatacamp.com

Get data
Watch znode for changes
Set data
Create children of a znode
List children of a znode
Check Status
Remove or Delete a znode

7)How Can We Create Znodes?

Answer)Create a znode with the given path. The flag argument specifies whether the created znode will be ephemeral, persistent, or sequential. By default, all znodes are persistent.

Ephemeral znodes (flag: e) will be automatically deleted when a session expires or when the client disconnects.

Sequential znodes guaranty that the znode path will be unique.

ZooKeeper ensemble will add sequence number along with 10 digit padding to the znode path. For example, the znode path /myapp will be converted to /myapp0000000001 and the next sequence number will be /myapp0000000002.
If no flags are specified, then the znode is considered as persistent.
create /path /data
To create a Sequential znode, add -s flag as shown below.
create -s /path /data
To create an Ephemeral Znode, add -e flag as shown below.
create -e /path /data

8)How Can We Create Children / Sub-znode?

Answer)Creating children is similar to creating new znodes. The only difference is that the path of the child znode will have the parent path as well.

create /parent/path/subnode/path /data

9)How Can We Remove A Znode?

Answer)Removes a specified znode and recursively all its children. This would happen only if such a znode is available.

rmr /path

10)What Are The Basics Of Zookeeper Api?

Answer)Application interacting with ZooKeeper ensemble is referred as ZooKeeper Client or simply Client. Znode is the core component of ZooKeeper ensemble and ZooKeeper API provides a small set of methods to manipulate all the details of znode with ZooKeeper ensemble. A client should follow the steps given below to have a clear and clean interaction with ZooKeeper ensemble.

Connect to the ZooKeeper ensemble. ZooKeeper ensemble assign a Session ID for the client.

Send heartbeats to the server periodically. Otherwise, the ZooKeeper ensemble expires the Session ID and the client needs to reconnect.

Get / Set the znodes as long as a session ID is active.

Disconnect from the ZooKeeper ensemble, once all the tasks are completed. If the client is inactive for a prolonged time, then the ZooKeeper ensemble will automatically disconnect the client.

11)Explain The Methods Of Zookeeperclass?

Answer)The central part of the ZooKeeper API is ZooKeeper class. It provides options to connect the ZooKeeper ensemble in its constructor and has the following methods -
connect - connect to the ZooKeeper ensemble
ZooKeeper(String connectionString, int sessionTimeout, Watcher watcher)
create - create a znode
create(String path, byte[] data, List acl, CreateMode createMode)
exists - check whether a znode exists and its information
exists(String path, boolean watcher)
getData - get data from a particular znode
getData(String path, Watcher watcher, Stat stat)
setData - set data in a particular znode
setData(String path, byte[] data, int version)
getChildren - get all sub-nodes available in a particular znode
getChildren(String path, Watcher watcher)
delete - get a particular znode and all its children
delete(String path, int version)
close - close a connection

12)Mention Some Instances Where Zookeeper Is Using?

Answer)Below are some of instances where Apache ZooKeeper is being utilized:

Apache Storm, being a real time stateless processing/computing framework, manages its state in ZooKeeper Service

Apache Kafka uses it for choosing leader node for the topic partitions

Apache YARN relies on it for the automatic failover of resource manager (master node)

www.smartdatacamp.com

Yahoo! utilties it as the coordination and failure recovery service for Yahoo! Message Broker, which is a highly scalable publish-subscribe system managing thousands of topics for replication and data delivery. It is used by the Fetching Service for Yahoo! crawler, where it also manages failure recovery.

13)Can Apache Kafka be used without Zookeeper?

Answer)It is not possible to use Apache Kafka without Zookeeper because if the Zookeeper is down Kafka cannot serve client request.

14)What is the role of Zookeeper in HBase architecture?

Answer)In HBase architecture, ZooKeeper is the monitoring server that provides different services like –tracking server failure and network partitions, maintaining the configuration information, establishing communication between the clients and region servers, usability of ephemeral nodes to identify the available servers in the cluster.

15)Explain about ZooKeeper in Kafka

Answer)Apache Kafka uses ZooKeeper to be a highly distributed and scalable system. Zookeeper is used by Kafka to store various configurations and use them across the hadoop cluster in a distributed manner. To achieve distributed-ness, configurations are distributed and replicated throughout the leader and follower nodes in the ZooKeeper ensemble. We cannot directly connect to Kafka by bye-passing ZooKeeper because if the ZooKeeper is down it will not be able to serve the client request.

16)Explain how Zookeeper works

Answer)ZooKeeper is referred to as the King of Coordination and distributed applications use ZooKeeper to store and facilitate important configuration information updates. ZooKeeper works by coordinating the processes of distributed applications. ZooKeeper is a robust replicated synchronization service with eventual consistency. A set of nodes is known as an ensemble and persisted data is distributed between multiple nodes.

3 or more independent servers collectively form a ZooKeeper cluster and elect a master. One client connects to any of the specific server and migrates if a particular node fails. The ensemble of ZooKeeper nodes is alive till the majority of nods are working. The master node in ZooKeeper is dynamically selected by the consensus within the ensemble so if the master node fails then the role of master node will migrate to another node which is selected dynamically. Writes are linear and reads are concurrent in ZooKeeper.

17)List some examples of Zookeeper use cases.

Answer)Found by Elastic uses Zookeeper comprehensively for resource allocation, leader election, high priority notifications and discovery. The entire service of Found built up of various systems that read and write to Zookeeper.
Apache Kafka that depends on ZooKeeper is used by LinkedIn
Storm that relies on ZooKeeper is used by popular companies like Groupon and Twitter.

18)How to use Apache Zookeeper command line interface?

Answer)ZooKeeper has a command line client support for interactive use. The command line interface of ZooKeeper is similar to the file and shell system of UNIX. Data in ZooKeeper is stored in a hierarchy of Znodes where each znode can contain data just similar to a file. Each znode can also have children just like directories in the UNIX file system.

Zookeeper-client command is used to launch the command line client. If the initial prompt is hidden by the log messages after entering the command, users can just hit ENTER to view the prompt.

19)What are the different types of Znodes?

Answer)There are 2 types of Znodes namely- Ephemeral and Sequential Znodes.
The Znodes that get destroyed as soon as the client that created it disconnects are referred to as Ephemeral Znodes.
Sequential Znode is the one in which sequential number is chosen by the ZooKeeper ensemble and is pre-fixed when the client assigns name to the znode.

20)What are watches?

Answer)Client disconnection might be troublesome problem especially when we need to keep a track on the state of Znodes at regular intervals. ZooKeeper has an event system referred to as watch which can be set on Znode to trigger an event whenever it is removed, altered or any new children are created below it.

21)What problems can be addressed by using Zookeeper?

Answer)In the development of distributed systems, creating own protocols for coordinating the hadoop cluster results in failure and frustration for the developers. The architecture of a distributed system can be prone to deadlocks, inconsistency and race conditions. This leads to various difficulties in making the hadoop cluster fast, reliable and scalable. To address all such

www.smartdatacamp.com

problems, Apache ZooKeeper can be used as a coordination service to write correct distributed applications without having to reinvent the wheel from the beginning.

22)How should I handle the CONNECTION_LOSS error?

Answer)CONNECTION_LOSS means the link between the client and server was broken. It doesn't necessarily mean that the request failed. If you are doing a create request and the link was broken after the request reached the server and before the response was returned, the create request will succeed. If the link was broken before the packet went onto the wire, the create request failed. Unfortunately, there is no way for the client library to know, so it returns CONNECTION_LOSS. The programmer must figure out if the request succeeded or needs to be retried. Usually this is done in an application specific way. Examples of success detection include checking for the presence of a file to be created or checking the value of a znode to be modified.

When a client (session) becomes partitioned from the ZK serving cluster it will begin searching the list of servers that were specified during session creation. Eventually, when connectivity between the client and at least one of the servers is re-established, the session will either again transition to the connected state (if reconnected within the session timeout value) or it will transition to the expired state (if reconnected after the session timeout). The ZK client library will handle reconnect for you automatically. In particular we have heuristics built into the client library to handle things like herd effect, etc. Only create a new session when you are notified of session expiration (mandatory).

23)How should I handle SESSION_EXPIRED?

Answer)SESSION_EXPIRED automatically closes the ZooKeeper handle. In a correctly operating cluster, you should never see SESSION_EXPIRED. It means that the client was partitioned off from the ZooKeeper service for more the the session timeout and ZooKeeper decided that the client died. Because the ZooKeeper service is ground truth, the client should consider itself dead and go into recovery. If the client is only reading state from ZooKeeper, recovery means just reconnecting. In more complex applications, recovery means recreating ephemeral nodes, vying for leadership roles, and reconstructing published state.

Library writers should be conscious of the severity of the expired state and not try to recover from it. Instead libraries should return a fatal error. Even if the library is simply reading from ZooKeeper, the user of the library may also be doing other things with ZooKeeper that requires more complex recovery.

Session expiration is managed by the ZooKeeper cluster itself, not by the client. When the ZK client establishes a session with the cluster it provides a timeout value. This value is used by the cluster to determine when the client's session expires. Expirations happens when the cluster does not hear from the client within the specified session timeout period (i.e. no heartbeat). At session expiration the cluster will delete any/all ephemeral nodes owned by that session and immediately notify any/all connected clients of the change (anyone watching those znodes). At this point the client of the expired session is still disconnected from the cluster, it will not be notified of the session expiration until/unless it is able to re-establish a connection to the cluster.

The client will stay in disconnected state until the TCP connection is re-established with the cluster, at which point the watcher of the expired session will receive the session expired notification.

24)Is there an easy way to expire a session for testing?

Answer)Yes, a ZooKeeper handle can take a session id and password. This constructor is used to recover a session after total application failure. For example, an application can connect to ZooKeeper, save the session id and password to a file, terminate, restart, read the session id and password, and reconnect to ZooKeeper without loosing the session and the corresponding ephemeral nodes. It is up to the programmer to ensure that the session id and password isn't passed around to multiple instances of an application, otherwise problems can result.

In the case of testing we want to cause a problem, so to explicitly expire a session an application connects to ZooKeeper, saves the session id and password, creates another ZooKeeper handle with that id and password, and then closes the new handle. Since both handles reference the same session, the close on second handle will invalidate the session causing a SESSION_EXPIRED on the first handle.

25)Why doesn't the NodeChildrenChanged and NodeDataChanged watch events return more information about the change?

Answer)When a ZooKeeper server generates the change events, it knows exactly what the change is. In our initial implementation of ZooKeeper we returned this information with the change event, but it turned out that it was impossible to use correctly. There may be a correct way to use it, but we have never seen a case of correct usage. The problem is that watches are used to find out about the latest change. (Otherwise, you would just do periodic gets.) The thing that most programmers seem to miss, when they ask for this feature, is that watches are one time triggers. Observe the following case of data change: a process does a getData on /a with watch set to true and gets v1, another process changes /a to v2 and shortly there after changes /a to v3. The first process would see that /a was changed to v2, but wouldn't know that /a is now /v3.

26)What are the options-process for upgrading ZooKeeper?

Answer)There are two primary ways of doing this; 1) full restart or 2) rolling restart.

In the full restart case you can stage your updated code/configuration/etc., stop all of the servers in the ensemble, switch code/configuration, and restart the ZooKeeper ensemble. If you do this programmatically (scripts typically, ie not by hand) the restart can be done on order of seconds. As a result the clients will lose connectivity to the ZooKeeper cluster during this time, however it looks to the clients just like a network partition. All existing client sessions are maintained and re-established as soon as the ZooKeeper ensemble comes back up. Obviously one drawback to this approach is that if you encounter any issues (it's always a good idea to test or stage these changes on a test harness) the cluster may be down for longer than expected.

The second option, preferable for many users, is to do a rolling restart. In this case you upgrade one server in the ZooKeeper ensemble at a time; bring down the server, upgrade the code/configuration/etc., then restart the server. The server will automatically rejoin the quorum, update it's internal state with the current ZK leader, and begin serving client sessions. As a result of doing a rolling restart, rather than a full restart, the administrator can monitor the ensemble as the upgrade progresses, perhaps rolling back if any issues are encountered.

27)What happens to ZK sessions while the cluster is down?

Answer)Imagine that a client is connected to ZK with a 5 second session timeout, and the administrator brings the entire ZK cluster down for an upgrade. The cluster is down for several minutes, and then is restarted.

In this scenario, the client is able to reconnect and refresh its session. Because session timeouts are tracked by the leader, the session starts counting down again with a fresh timeout when the cluster is restarted. So, as long as the client connects within the first 5 seconds after a leader is elected, it will reconnect without an expiration, and any ephemeral nodes it had prior to the downtime will be maintained.

The same behavior is exhibited when the leader crashes and a new one is elected. In the limit, if the leader is flip-flopping back and forth quickly, sessions will never expire since their timers are getting constantly reset.

www.smartdatacamp.com

Apache Yarn

Apache Yarn a platform responsible for managing computing resources in clusters and using them for scheduling users applications

1)What is Apache Hadoop YARN?

Answer)YARN stands for 'Yet Another Resource Negtiator'.YARN is a powerful and efficient feature rolled out as a part of Hadoop 2.0.YARN is a large scale distributed system for running big data applications.

2)What are the core concepts in YARN?

Answer)Resource Manager: As equivalent to JobTracker
Node Manager: As equivalent to TaskTracker
Application Manager: As equivalent to Jobs.
Containers: As quivalent to slots
YARN child: After submitting the application, dynamically application master launch YARN child to do the MapReduce tasks.

3)Is YARN a replacement of Hadoop MapReduce?

Answer)YARN is not a replacement of Hadoop but it is a more powerful and efficient technology that supports MapReduce and is also referred to as Hadoop 2.0 or MapReduce 2.

4)What are the additional benefits YARN brings in to Hadoop?

Answer)Effective utilization of the resources as multiple applications can be run in YARN all sharing a common resource.YARN is backward compatible so all the existing MapReduce jobs.Using YARN, one can even run applications that are not based on the MaReduce model

5)How can native libraries be included in YARN jobs?

Answer)There are two ways to include native libraries in YARN jobs-
1) By setting the -Djava.library.path on the command line
2) By setting the LD_LIBRARY_PATH in the .bashrc filewer

www.smartdatacamp.com

6)What is a container in YARN? Is it same as the child JVM in which the tasks on the nodemanager run or is it different?

Answer)It represents a resource (memory) on a single node at a given cluster.
A container is
supervised by the node manager
scheduled by the resource manager
One MR task runs in such container(s).

Apache Oozie

Apache Oozie is a server-based workflow scheduling system to manage Hadoop jobs.Workflows in Oozie are defined as a collection of control flow and action nodes in a directed acyclic graph. Control flow nodes define the beginning and the end of a workflow (start, end, and failure nodes) as well as a mechanism to control the workflow execution path (decision, fork, and join nodes).

1)What is Oozie?

Answer)Oozie is a workflow scheduler for Hadoop Oozie allows a user to create Directed Acyclic Graphs of workflows and these can be ran in parallel and sequential in Hadoop.It can also run plain java classes, Pig workflows and interact with the HDFS .It can run jobs sequentially and in parallel.

2)Why use oozie instead of just cascading a jobs one after another?

Answer)Major Flexibility :Start ,stop ,re-run and suspend
Oozie allows us to restart from failure

3)How to make a workflow?

Answer)First make a Hadoop job and make sure that it works Make a jar out of classes and then make a workflow.xml file and copy all of the job configuration properties in to the xml file.
Input files
Output files
Input readers and writers
mappers and reducers
job specific arguments
job.properties

4)What are the properties that we have to mention in .Properties?

Answer)Name Node
Job Tracker
Oozie.wf.application.path
Lib Path
Jar Path

www.smartdatacamp.com

5)What is application pipeline in Oozie?

Answer)It is necessary to connect workflow jobs that run regularly, but at different time intervals. The outputs of multiple subsequent runs of a workflow become the input to the next workflow. Chaining together these workflows result it is referred as a data application pipeline.

6)How to run Oozie?

Answer)$ oozie job -oozie http://172.20.95.107:11000(oozie server node)/oozie -config job.properties -run

This will give the job id.

To know the status: $ oozie job -oozie http://172.20.95.107:11000(oozie server node)/oozie -info job id

7)What are all the actions can be performed in Oozie?

Answer)Email Action
Hive Action
Shell Action
Ssh Action
Sqoop Action
Writing a custom Action Executor

8)Why we use Fork and Join nodes of oozie?

Answer)A fork node splits one path of execution into multiple concurrent paths of execution.

A join node waits until every concurrent execution path of a previous fork node arrives to it.

The fork and join nodes must be used in pairs. The join node assumes concurrent execution paths are children of the same fork node.

Apache CouchDB

Apache CouchDB is open source database software that focuses on ease of use and having a scalable architecture. It has a document-oriented NoSQL database architecture and is implemented in the concurrency-oriented language Erlang; it uses JSON to store data, JavaScript as its query language using MapReduce, and HTTP for an API.

1)What Language Is Couchdb Written In ?

Answer)Erlang, a concurrent, functional programming language with an emphasis on fault tolerance.
Early work on CouchDB was started in C++ but was replaced by Erlang OTP platform. Erlang has so far proven an excellent match for this project.
CouchDB's default view server uses Mozilla's Spidermonkey JavaScript library which is written in C. It also supports easy integration of view servers written in any language

2)Why Does Couchdb Not Use Mnesia?

Answer)The first is a storage limitation of 2 Giga bytes per file.
The second is that it requires a validation and fix up cycle after a crash or power failure, so even if the size limitation is lifted, the fix up time on large files is prohibitive.
Mnesia replication is suitable for clustering, but not disconnected, distributed edits. Most of the cool features of Mnesia aren't really useful for CouchDB.
Also Mnesia isn't really a general-purpose, large scale database. It works best as a configuration type database, the type where the data isn't central to the function of the application, but is necessary for the normal operation of it. Think things like network routers, HTTP proxies and LDAP directories, things that need to be updated, configured and reconfigured often, but that configuration data is rarely very large.

3)How Do I Use Transactions With Couchdb?

Answer)CouchDB uses an Optimistic concurrency model. In the simplest terms, this just means that you send a document version along with your update, and CouchDB rejects the change if the current document version doesn't match what you've sent.
You can re-frame many normal transaction based scenarios for CouchDB. You do need to sort of throw out your RDBMS domain knowledge when learning CouchDB, though.
It's helpful to approach problems from a higher level, rather than attempting to mold Couch to a SQL based world.

4)How Do You Compare Mongodb, Couchdb And Couchbase?

www.smartdatacamp.com

Answer)MongoDB and CouchDB are document oriented database.
MongoDB and CouchDB are the most typical representative of the open source NoSQL database. They have nothing in common other than are stored in the document outside.
MongoDB and CouchDB, the data model interface, object storage and replication methods have many different.

5)How Is Pouchdb Different From Couchdb?

Answer)PouchDB is also a CouchDB client, and you should be able to switch between a local database or an online CouchDB instance without changing any of your application's code. However, there are some minor differences to note:
View Collation – CouchDB uses ICU to order keys in a view query; in PouchDB they are ASCII ordered.
View Offset – CouchDB returns an offset property in the view results. In PouchDB, offset just mirrors the skip parameter rather than returning a true offset.

6)So Is Couchdb Now Going To Written In Java?

Answer)Erlang is a great fit for CouchDB and I have absolutely no plans to move the project off its Erlang base. IBM/Apache's only concerns are we remove license incompatible 3rd party source code bundled with the project, a fundamental requirement for any Apache project. So some things may have to replaced in the source code (possibly Mozilla Spidermonkey), but the core Erlang code stays.
An important goal is to keep interfaces in CouchDB simple enough that creating compatible implementations on other platforms is feasible. CouchDB has already inspired the database projects RDDB and Basura. Like SQL databases, I think CouchDB needs competition and a ecosystem to be viable long term. So Java or C++ versions might be created and I would be delighted to see them, but it likely won't be me who does it.

7)What Does Ibm's Involvement Mean For Couchdb And The Community?

Answer)The main consequences of IBM's involvement are:
The code is now being Apache licensed, instead of GPL.
Damien is going to be contributing much more time

8)Mention The Main Features Of Couchdb?

Answer)JSON Documents – Everything stored in CouchDB boils down to a JSON document.
RESTful Interface – From creation to replication to data insertion, every management and data task in CouchDB can be done via HTTP.

N-Master Replication – You can make use of an unlimited amount of 'masters', making for some very interesting replication topologies.
Built for Offline – CouchDB can replicate to devices (like Android phones) that can go offline and handle data sync for you when the device is back online.
Replication Filters – You can filter precisely the data you wish to replicate to different nodes.

9)What Is The Use Of Couchdb?

Answer)CouchDB allows you to write a client side application that talks directly to the Couch without the need for a server side middle layer, significantly reducing development time. With CouchDB, you can easily handle demand by adding more replication nodes with ease. CouchDB allows you to replicate the database to your client and with filters you could even replicate that specific user's data.
Having the database stored locally means your client side application can run with almost no latency. CouchDB will handle the replication to the cloud for you. Your users could access their invoices on their mobile phone and make changes with no noticeable latency, all whilst being offline. When a connection is present and usable, CouchDB will automatically replicate those changes to your cloud CouchDB.
CouchDB is a database designed to run on the internet of today for today's desktop-like applications and the connected devices through which we access the internet

10)How Much Stuff Can Be Stored In Couchdb?

Answer)For node partitioning, basically unlimited. The practical scaling limits for a single database instance, are not yet known.

11)What Is Couchdb Kit?

Answer)The Couchdb Kit is used to provide a structure for your Python applications to manage and access Couchdb. This kit provides full featured and easy client to manage and access Couchdb. It helps you to maintain databases, to view access, Couchdb server and doc managements. Mostly python objects are reflected by the objects for convenience. The Database and server objects are used easily as using a dict.

12)Can Views Update Documents Or Databases?

Answer)No. Views are always readonly to databases and their documents.

13)Where Are The Couchdb Logfiles Located?

www.smartdatacamp.com

Answer)For a default linux/unix installation the logfiles are located here:
/usr/local/var/log/couchdb/couch.log
This is set in the default.ini file located here:
/etc/couchdb/default.ini
If you've installed from source and are running couchdb in dev mode the logfiles are located here:
YOURCOUCHDBSOURCEDIRECTORY/tmp/log/couch.log

14)What Does Couch Mean?

Answer)It's an acronym, Cluster Of Unreliable Commodity Hardware. This is a statement of Couch's long term goals of massive scalability and high reliability on fault prone hardware. The distributed nature and flat address space of the database will enable node partitioning for storage scalability (with a map/reduce style query facility) and clustering for reliability and fault tolerance.

15)Is Couchdb Ready For Production?

Answer)Yes. There are many companies using CouchDB.

16)What Platforms Are Supported?

Answer)Most POSIX systems, this includes GNU/Linux and OS X.
Windows is not officially supported but it should work

17)How Do I Do Sequences?

Answer)With replication sequences are hard to realize. Sequences are often used to ensure unique identifiers for each row in a database table. CouchDB generates unique ids from its own and you can specify your own as well, so you don't really need a sequence here. If you use a sequence for something else, you might find a way to express in CouchDB in another way.

18)How Do I Use Replication?

Answer)POST /_replicate with a post body of
{"source":"$source_database"
,
"target":"$target_database"}
Where $source_database and $target_database can be the names of local database or full URIs of remote databases. Both databases need to be created before they can be replicated from or to.

19)How Do I Review Conflicts Occurred During Replication?

Answer)Use a view like this:
map: function(doc) {if(doc._conflicts){emit(null,null);}}

20)How Can I Spread Load Across Multiple Nodes?

Answer)Using an http proxy like nginx, you can load balance GETs across nodes, and direct all POSTs, PUTs and DELETEs to a master node. CouchDB's triggered replication facility can keep multiple read-only servers in sync with a single master server, so by replicating from master > slaves on a regular basis, you can keep your content up to date.

21)Can I Talk To Couchdb Without Going Through The Http Api?

Answer)CouchDB's data model and internal API map the REST/HTTP model so well that any other API would basically reinvent some flavor of HTTP. However, there is a plan to refractor CouchDB's internals so as to provide a documented Erlang API.

22)Erlang Has Been Slow To Adopt Unicode. Is Unicode Or Utf8 A Problem With Couchdb?

Answer)CouchDB uses Erlang binaries internally. All data coming to CouchDB must be UTF8 encoded.

23)How Fast Are Couchdb Views?

Answer)It would be quite hard to give out any numbers that make much sense. From the architecture point of view, a view on a table is much like a (multicolumn) index on a table in an RDBMS that just performs a quick lookup. So this theoretically should be pretty quick. The major advantage of the architecture is, however, that it is designed for high traffic. No locking occurs in the storage module (MVCC and all that) allowing any number of parallel readers as well as serialized writes. With replication, you can even set up multiple machines for a horizontal scaleout and data partitioning (in the future) will let you cope with huge volumes of data

24)Is it possible to communicate to CouchDB without going through HTTP/ API?

Answer)CouchDB's data model and internal API map the REST/HTTP model in a very simple way that any other API would basically inherit some features of HTTP. However, there is a plan to refractor CouchDB's internals so as to provide a documented Erlang API.

www.smartdatacamp.com

25)What Platforms are Supported?

Answer)Most POSIX systems,this includes GNU/Linux and OS X.
Windows is not officially supported but it should work.

26)My database will require an unbounded number of deletes, what can I do?

Answer)If there's a strong correlation between time (or some other regular monotonically increasing event) and document deletion, a DB setup can be used like the following:
Assume that the past 30 days of logs are needed, anything older can be deleted.
Set up DB logs_2011_08.
Replicate logs_2011_08 to logs_2011_09, filtered on logs from 2011_08 only.
During August, read/write to logs_2011_08.
When September starts, create logs_2011_10.
Replicate logs_2011_09 to logs_2011_10, filtered on logs from 2011_09 only.
During September, read/write to logs_2011_09.
Logs from August will be present in logs_2011_09 due to the replication, but not in logs_2011_10.
The entire logs_2011_08 DB can be removed.

27)How do I backup CouchDB? What data recovery strategies exist?

Answer)While CouchDB is a very reliable database, a careful engineer will always ask "What happens when something goes wrong?". Let's say your server has an unrecoverable crash and you lose all data... or maybe a hacker finds your top secret credentials and deletes your data... or maybe an undiscovered bug causes data corruption after an event... or maybe there is a logic error in your application code that accesses your database. Ideally we try to avoid these situations by preparing for the worst and hoping they never occur, but bad things do happen and we should be ready to react when they do. There are a few traditional data backup strategies for CouchDB: Replication Database file backup Filesystem snapshots Replication Based Backup CouchDB is well known for its push and pull replication functionality. Any CouchDB database can replicate to any other if it has HTTP access and the proper credentials. Database File Backup Under the hood, CouchDB stores databases and indexes as files in the underlying filesystem. Using a common command line back up tool, like rsync, we can perform incremental backups triggered by cron. Filesystem/VM Snapshots Most VM's and newer filesystems have snapshot capabilities to allow roll backs to preserve data.

28)How Do I Configure SSL (HTTPS) in CouchDB?

Answer)Secure Socket Layer (SSL) is used in conjunction with HTTP to secure web traffic. The resulting protocol is known as HTTPS. In order to utilize SSL, you must generate a key and cert. Additionally, if you want your web traffic to be safely accepted by most web browsers, you will

need the cert to be signed by a CA (Certificate Authority). Otherwise, if you bypass the CA, you have the option of self signing your certificate. Production Security Apache CouchDB leverages Erlang/OTP's SSL, which is usually linked against a system-provided OpenSSL installation. The security, performance & compatibility with other browsers and operating systems therefore varies heavily depending on how the underlying OpenSSL library was set up. It is strongly recommended that for production deployments, a dedicated well-known SSL/TLS terminator is used instead. There is nothing fundamentally wrong with Erlang's crypto libraries, however a dedicated TLS application is generally a better choice, and allows tuning and configuring your TLS settings directly rather than relying on whatever Erlang/OTP release is provided by your operating system. Key & CSR Procedure using OpenSSL OpenSSL is an open source SSL utility and library. It comes standard with many UNIX/LINUX distributions. We will use OpenSSL to generate our private key and generate our certificate signing request (CSR).

28)What are the consequences of having a high ratio of 'deleted' to 'active' documents?

Answer)Every document that is deleted is replaced with small amount of metadata called a tombstone which is used for conflict resolution during replication (a tombstone is also created for each document that is in a batch delete operation). Although tombstone documents contain only a small amount of metadata, having lots of tombstone documents will have an impact on the size of used storage. Tombstone documents still show up in _changes so require processing for replication and when building views. Compaction time is proportional to the ratio of deleted documents to the total document count.

29)Deleted documents have an overhead in CouchDB because a tombstone document exists for each deleted document. One consequence of tombstone documents is that compaction gets slower over time. Three options for purging tombstone documents from a CouchDB are: Create a new database for every N time period (and delete that database when the period expires) Filtered replication Do nothing How can I choose which option is the most suitable?

Answer)Each approach is described below. Note that you may need to use a combination of both approaches in your application. Alternatively, you may find through testing that your tombstone documents don't add significant overhead and can just be left as is. Create a new database for every N time period When to use this approach? This approach works best when you know the expiry date of a document at the time when the document is first saved. How does it work? Each document to be saved that has a known expiry date will be stored in a database that will get dropped when its expiry date has been reached. When the document is being saved, if the database doesn't already exist then a new database must be created. The rationale of this approach is that dropping a database is an in-expensive operation and does not leave tombstone documents on disk. Gotchas It is not possible to query across database in Cloudant/CouchDB. Cross database queries will need to be performed in the application itself. This will be an issue if the cross database queries require aggregating lots of data. Filtered replication When to use it This approach works best when you don't know the expiry date of a document at the time when the document is first saved, or if you would have to perform cross database queries that would

www.smartdatacamp.com

involve moving lots of data to the application so that it can be aggregated. How does it works? This approach relies on creating a new database at an opportune time (NOTE 1) and by replicating all documents to it except for the tombstone documents. A validate_doc_update (VDU) function is used so that deleted documents with no existing entry in the target database are rejected. When replication is complete (or acceptably up-to-date if using continuous replication), switch your application to use the new database and delete the old one. There is currently no way to rename databases but you could use a virtual host which points to the "current" database. An example of such a VDU function is below function (newDoc, oldDoc, userCtx) { // any update to an existing doc is OK if(oldDoc) { return; } // reject tombstones for docs we don't know about if(newDoc["_deleted"]) { throw({forbidden : "We're rejecting tombstones for unknown docs"}) } }

30)My filtered replication takes forever. Why is my filtered replication so slow?

Answer)Filtered replications works slow because for each fetched document runs complex logic to decision: to replicate it or not.

Apache Accumulo

Apache Accumulo is a highly scalable structured store based on Google's BigTable. Accumulo is written in Java and operates over the Hadoop Distributed File System (HDFS), which is part of the popular Apache Hadoop project. Accumulo supports efficient storage and retrieval of structured data, including queries for ranges, and provides support for using Accumulo tables as input and output for MapReduce jobs. Accumulo features automatic load-balancing and partitioning, data compression and fine-grained security labels.

1) How to remove instance of accumulo? We have created a instance while initializing accumulo by calling accumulo init But now i want to remove that instance and as well i want to create a new instance. Can any one help to do this?

Answer)Remove the directory specified by the instance.dfs.dir property in $ACCUMULO_HOME/conf/accumulo-site.xml from HDFS.
If you did not specify an instance.dfs.dir in accumulo-site.xml, the default is "/accumulo".
You should then be able to call accumulo init with success.

2) How are the tablets mapped to a Datanode or HDFS block? Obviously, One tablet is split into multiple HDFS blocks (8 in this case) so would they be stored on the same or different datanode(s) or does it not matter?

Answer) Tablets are stored in blocks like all other files in HDFS. You will typically see all blocks for a single file on at least one data node (this isn't always the case, but seems to mostly hold true when i've looked at block locations for larger files)

3) In the example above, would all data about RowC (or A or B) go onto the same HDFS block or different HDFS blocks?

Answer) Depends on the block size for your tablets (dfs.block.size or if configured the Accumulo property table.file.blocksize). If the block size is the same size as the tablet size, then obviously they will be in the same HDFS block. Otherwise if the block size is smaller than the tablet size, then it's pot luck as to whether they are in the same block or not.

4) When executing a map reduce job how many mappers would I get? (one per hdfs block? or per tablet? or per server?)

www.smartdatacamp.com

Answer) This depends on the ranges you give InputFormatBase.setRanges(Configuration, Collection<Ranges>).

If you scan the entire table (-inf -> +inf), then you'll get a number of mappers equal to the number of tablets (caveated by disableAutoAdjustRanges). If you define specific ranges, you'll get a different behavior depending on whether you've called InputFormatBase.disableAutoAdjustRanges(Configuration) or not:

If you have called this method then you'll get one mapper per range defined. Importantly, if you have a range that starts in one tablet and ends in another, you'll get one mapper to process that entire range

If you don't call this method, and you have a range that spans over tablets, then you'll get one mapper for each tablet the range covers

5) How do I create a Spark RDD from Accumulo?

Answer) Generally with custom Hadoop InputFormats, the information is specified using a JobConf. As @Sietse pointed out there are some static methods on the AccumuloInputFormat that you can use to configure the JobConf. In this case I think what you would want to do is:

```
val jobConf = new JobConf() // Create a job conf
// Configure the job conf with our accumulo properties
AccumuloInputFormat.setConnectorInfo(jobConf, principal, token)
AccumuloInputFormat.setScanAuthorizations(jobConf, authorizations)
val clientConfig = new
ClientConfiguration().withInstance(instanceName).withZkHosts(zooKeepers)
AccumuloInputFormat.setZooKeeperInstance(jobConf, clientConfig)
AccumuloInputFormat.setInputTableName(jobConf, tableName)
// Create an RDD using the jobConf
val rdd2 = sc.newAPIHadoopRDD(jobConf,
classOf[org.apache.accumulo.core.client.mapreduce.AccumuloInputFormat],
classOf[org.apache.accumulo.core.data.Key],
classOf[org.apache.accumulo.core.data.Value]
)
```

6) How to filter Scan on Accumulo using RegEx?

Answer)The Filter class lays the framework for the functionality you want. To create a custom filter, you need to extend Filter and implement the accept(Key k, Value v) method. If you are only looking to filter based on regular expressions, you can avoid writing your own filter by using RegExFilter.

Using a RegExFilter is straightforward. Here is an example:

```
//first connect to Accumulo
```

```
ZooKeeperInstance inst = new ZooKeeperInstance(instanceName, zooServers);
Connector connect = inst.getConnector(user, password);

//initialize a scanner
Scanner scan = connect.createScanner(myTableName, myAuthorizations);

//to use a filter, which is an iterator, you must create an IteratorSetting
//specifying which iterator class you are using
IteratorSetting iter = new IteratorSetting(15, "myFilter", RegExFilter.class);
//next set the regular expressions to match. Here, I want all key/value pairs in
//which the column family begins with "J"
String rowRegex = null;
String colfRegex = "J.*";
String colqRegex = null;
String valueRegex = null;
boolean orFields = false;
RegExFilter.setRegexs(iter, rowRegex, colfRegex, colqRegex, valueRegex, orFields);
//now add the iterator to the scanner, and you're all set
scan.addScanIterator(iter);
```

The first two parameters of the iteratorSetting constructor (priority and name) are not relevant in this case. Once you've added the above code, iterating through the scanner will only return key/value pairs that match the regex parameters.

7) Connecting to Accumulo inside a Mapper using Kerberos

Answer) The provided AccumuloInputFormat and AccumuloOutputFormat have a method to set the token in the job configuration with the Accumulo*putFormat.setConnectorInfo(job, principle, token). You can also serialize the token in a file in HDFS, using the AuthenticationTokenSerializer and use the version of the setConnectorInfo method which accepts a file name.

If a KerberosToken is passed in, the job will create a DelegationToken to use, and if a DelegationToken is passed in, it will just use that.

The provided AccumuloInputFormat should handle its own scanner, so normally, you shouldn't have to do that in your Mapper if you've set the configuration properly. However, if you're doing secondary scanning (for something like a join) inside your Mapper, you can inspect the provided AccumuloInputFormat's RecordReader source code for an example of how to retrieve the configuration and construct a Scanner.

8)How to get count for database query in Accumulo

Answer)Accumulo is a lower-level application than a traditional RDBMS. It is based on Google's Big Table and not like a relational database. It's more accurately described as a massive parallel sorted map than a database.

It is designed to do different kinds of tasks than a relational database, and its focus is on big data.

To achieve the equivalent of the MongoDB feature you mentioned in Accumulo (to get a count of the size of an arbitrary query's result set), you can write a server-side Iterator which returns counts from each server, which can be summed on the client side to get a total. If you can anticipate your queries, you can also create an index which keeps track of counts during the ingest of your data.

Creating custom Iterators is an advanced activity. Typically, there are important trade-offs (time/space/consistency/convenience) to implementing something as seemingly simple as a count of a result set, so proceed with caution. I would recommend consulting the user mailing list for information and advice.

9)How do you use "Range" to Scan an entire table in accumulo

Answer)This is the same thing that the previous answer is saying, but I thought it might help to show a line of code.

If you have a scanner, cleverly named 'scanner', you can use the setRange() method to set the range on the scanner. Because the default range is (-inf, +inf), passing setRange a newly created range object will give your scanner, with a range of (-inf, +inf), the ability to scan the entire table.

The sample code looks like:

scanner.setRange(new Range());

10)What CAP-Type does Apache Accumulo have?

Answer) Apache Accumulo is based on the Google BigTable paper, and shares a lot of similarities with Apache HBase. All three of these systems are intended to be CP, where nodes will simply go down rather than serve inconsistent data.

11)How do I set an environment variable in a YARN Spark job?

Answer)So I discovered the answer to this while writing the question (sorry, reputation seekers). The problem is that CDH5 uses Spark 1.0.0, and that I was running the job via YARN. Apparently, YARN mode does not pay any attention to the executor environment and instead uses the environment variable SPARK_YARN_USER_ENV to control its environment. So ensuring SPARK_YARN_USER_ENV contains ACCUMULO_CONF_DIR=/etc/accumulo/conf works, and makes ACCUMULO_CONF_DIR visible in the environment at the indicated point in the question's source example.

12) Does Accumulo actually need all Zookeeper servers listed?

Answer) ZooKeeper servers operate as a coordinated group, where the group as a whole determines the value of a field at any given time, based on consensus among the servers. If you have a 5-node ZooKeeper instance running, all 5 server names are relevant. You should not simply treat them as 5 redundant 1-node instances. Accumulo, and other ZooKeeper clients, actually use all of the servers listed.

www.smartdatacamp.com

Apache Airavata

Apache Airavata is a framework that supports execution and management of computational scientific applications and workflows in grid-based systems, remote clusters and cloud-based systems. Airavata's main focus is on submitting and managing applications and workflows in grid based systems. Airavata's architecture is extensible to support for other underlying resources as well.

1) I have setup my own gateway and Airavata. When I log into the gateway I cannot create Compute resources. What should I do?

Answer) In your pga_config.php (in folder .../testdrive/app/config) under heading 'Portal Related Configurations' set 'super-admin-portal' => false, to true.

2)I don't get notifications when users create new accounts in my gateway. Why?

Answer: That's because you have not defined an email address in
'admin-emails' => ['xxx@xxx.com','yyy@yyy.com']. Here you can add one or many.

3)I am not receiving email notifications from compute resources for job status changes. What should I do?

Answer: In airavata-server.properties please locate and set your email account information.
email.based.monitor.host=imap.gmail.com
email.based.monitor.address=airavata-user@kuytje.nl
email.based.monitor.password=zzzz
email.based.monitor.folder.name=INBOX
email.based.monitor.store.protocol=imaps (either imaps or pop3)

4)In my Airavata log I have error messages like
ERROR org.apache.airavata.api.server.handler.AiravataServerHandler - Error occurred while retrieving SSH public keys for gateway
ERROR org.apache.airavata.credential.store.server.CredentialStoreServerHandler - Error occurred while retrieving credentials
What should I do?

Answer: This could be due to missing tables in your credential store database. Check whether CREDENTIALS and COMMUNITY_USER tables exits. If not create then using
CREATE TABLE COMMUNITY_USER
(

```
GATEWAY_ID VARCHAR(256) NOT NULL,
COMMUNITY_USER_NAME VARCHAR(256) NOT NULL,
TOKEN_ID VARCHAR(256) NOT NULL,
COMMUNITY_USER_EMAIL VARCHAR(256) NOT NULL,
PRIMARY KEY (GATEWAY_ID, COMMUNITY_USER_NAME, TOKEN_ID)
);
CREATE TABLE CREDENTIALS
(
GATEWAY_ID VARCHAR(256) NOT NULL,
TOKEN_ID VARCHAR(256) NOT NULL,
CREDENTIAL BLOB NOT NULL,
PORTAL_USER_ID VARCHAR(256) NOT NULL,
TIME_PERSISTED TIMESTAMP DEFAULT NOW() ON UPDATE NOW(),
PRIMARY KEY (GATEWAY_ID, TOKEN_ID)
);
```

5)I cannot login to my Compute Resource and launch jobs from Airavata using the SSH key I generated. What should I do?

Answer: Steps to use generated SSH key
- Generate SSH key + token using Credential Store
- Add the generated token to your resource through PGA --> Admin Dashboard --> Gateway Preferences
- Add the generated SSH key

6)When installing PGA in MAC i got below error after updating the composer.
- Error
Mcrypt PHP extension required.
Script php artisan clear-compiled handling the post-update-cmd event returned with an error
[RuntimeException]

Answer: Install mcrypt installation;

7)After following the required steps only the home page is working and some images are not shown properly.

Answer: If you are facing this behavior first check whether you have enabled mod_rewrite module in apache webserver.
And also check whether you have set AllowOverride All in the Vhost configuration file in apache web server.
(e.g file location is /etc/apache2/sites-available/default and there should be two places where you want to change)
ServerAdmin webmaster@dummy-host.example.com

www.smartdatacamp.com

DocumentRoot /var/www/html/portal/public
ServerName pga.example.com
AllowOverride all
ErrorLog logs/pga_error_log
CustomLog logs/pga--access_log common

8)I get the Error message Permission Denied to app/storage directory.

Answer: Execute the following command and grant all permissions;
sudo chmod -R 777 app/storage

9)In Ubuntu environment when executing sudo composer update it fails with message "Mcrypt PHP extension required".

Answer: To fix this install PHP mcrypt extension by following the below steps;
sudo apt-get install php5-mcrypt
Locate mcrypt.so ,to get its location Locate mcrypt.ini and open the mcrypt.ini file
sudo pico /etc/php5/mods-available/mcrypt.ini
Change the at line a extension= eg:/usr/lib/php5/20121212/mcrypt.so Save changes. Execute the command:
sudo php5enmod mcrypt
Now restart the apache server again and test PGA web-interface

10)When tried to login or create a new user account an Error is thrown which is similar to PHP Fatal error: SOAP-ERROR: Parsing WSDL: Couldn't load from...

Answer: If you face this kind of an error first check whether you have enabled PHP SOAP and OpenSSL extensions. If even after enabling them the issue is still occurring try updating the PHP OpenSSL extension. (Using command like yum update openssl)

11)If you are seeing an error similar to following in your airavata.log

Error: ERROR org.apache.airavata.registry.core.app.catalog.impl.StorageResourceImpl - Error while retrieving storage resource... javax.persistence.NoResultException: Query "SELECT p FROM StorageResource p WHERE p.storageResourceId =:param0" selected no result, but expected unique result.

Answer: Add storage resource ID in to the pga_config.php in your gateway code/app/config directory.

12)I am getting error

2016-05-19 16:17:08,225 [main] ERROR org.apache.airavata.server.ServerMain - Server Start Error: java.lang.RuntimeException: Failed to create database connection pool.

What should i do?

Answer: Airavata cannot create database connection because the mysql jar is not existing. Please follow step 8 of documentation in Installation --> Airavata --> Airavata Installation

13)What each application input property mean?

Answer:
- Name:
Identifier of the application input
- Value:
This could be a STRING value or it can also be used to set input file name.
- Type:
Input type, List contain STRING, INTEGER, FLOAT, URI, STDOUT and STDERR
- Application Arguments:
These are are the characters you would want on commandline in job execution for each input file or character input.
- Standard Input:
Futuristic property and not in real use at the moment
- User Friendly Description:
A description about the input to the gateway user. This will be displayed for users at experiment creation.
- Input Order:
this is a number field. This will be the order inputs displayed in experiment creation.
- Data is Staged:
- Is the Input Required:
Futuristic property and not in real use at the moment. Whether set to true or false all inputs are currently treated as mandatory.
- Required in Commandline:
When this is set to true the arguments and the input file or the value will be available in job execution commandline in job script.
- Meta Data:

14)Application Output Properties

Answer: Add Application Output
- Name:
This is the label for the output.
- Value:
This would be the actual name of the output airavata brings back to the PGA.
- Type:
Tye of the output. This is mostly depended on the application. To troubleshoot for almost all applications define STDOUT and STDERR

www.smartdatacamp.com

- Application Argument:
This would be arguments for outputs need to be in commandline.
- Data Movement:
Futuristic property and not in real use at the moment. Whether set to true or false all outputs are currently brought back to PGA.
- Is the Output required?:
- Required on command line?:
When this is set to true the arguments and the output file or the value will be available in job execution commandline in job script.
- Location:
- Search Query:

15)How to make input file available as an executable?

Answer:
- Input files defined are copied to the experiment working directory.
- Input files will be available in commandline when 'Required on Commandline' = true
- To add a commandline argument for a input file add 'Application Argument' for each input file. This will also define the order of files in commandline.

16)In Application Interface what is the use of 'Enable Optional File Inputs'

Answer: - By setting 'Enable Optional File Inputs' = true user can add none or many input files at experiment creation.
- In Airavata any input file required for the application to execute need to be defined as a separate input.
- When inputs are defined they are treated as 'Mandatory' inputs.

17)Where do I add remote resource execution commands?

Answer: In Admin Dashboard --> Application Deployment
- Add module load commands
- Pre and post job commands
- Environment variables

18)What is the Project?

Answer:
- Project is simply a collection of experiments.
- When creating an experiment it will be under the project you select.
- Projects can be shared with fellow gateway users.
- When the project is shared all experiments in the project will be shared as well.

19)Where can I get test inputs?

Answer: If you need test inputs try downloading from Test Input Files

20)Do I need to provide values for wall-time, node count and CPU count? OR can I go ahead with the given default values?

Answer: Default values are given to make life little easy for users. But depending on the job you should change the values.
E.g.: If you need only two or less cores than 16 better to change. If you need more wall-time change it, etc....

21)What can I do with email Notifications in experiment?

Answer: Submitting a job from PGA does not guarantee job getting executed in the remote resource right away. If you add your email address, when the job starts and completes an email will be sent to you from the remote resource.

Why 'Save' and 'Save and Launch'?

Answer: User has the option of either create and 'Save' the experiment for later launch at remote resource
Or
can directly 'Save and Launch' at once.

23)How do I monitor progress?

Answer: When the experiment is launched navigate to Experiment Summary Page. There the experiment and job status will be present. You can monitor the experiment and job status changes. Refresh the status using 'Refresh' icon on top of the summary page.

24)How do I view/download my outputs?

Answer: Same way you monitor the experiment progress!
For each experiment inputs can be downloaded from
- Experiment Summary page
- From 'Storage' at the top menu. Here you need to know the Project the experiment belongs to. Once you locate the project you can browse for your experiment.

www.smartdatacamp.com

25)I want to bring back all the outputs generated by my job. How?

Answer: You need to request your gateway admin to simply set Archive to 'true' in respective application interface in Admin Dashboard. Then all the files in the working directory will be brought back to PGA. This flag is set at gateway level not at individual user level.

26)I want to cancel my experiment. How?

Answer: Navigate to experiment Summary and click 'Cancel' button. Experiments only in operation (LAUNCHED, EXECUTING experiment statuses) can be cancelled.

27)When I cancel do I still receive any outputs generated?

Answer: Files will be not transferred from the remote resource if the experiment is cancelled. However, if the output files were transferred to PGA data directories prior to the cancel request then they will be displayed.

28)How can I run same application with different different inputs?

Answer: Simply clone an existing experiment and change the input files and launch.

29)I want to change the wall-time of my experiment before launch. How can I do?

Answer: Experiments in CREATED state can be modified through Experiment Summary page. Click 'Edit' change values and Save or Save and Launch.

www.smartdatacamp.com

Apache Ambari

The Apache Ambari project is aimed at making Hadoop management simpler by developing software for provisioning, managing, and monitoring Apache Hadoop clusters. Ambari provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs.
Ambari enables System Administrators to:
Provision a Hadoop Cluster
Ambari provides a step-by-step wizard for installing Hadoop services across any number of hosts. Ambari handles configuration of Hadoop services for the cluster.
Manage a Hadoop Cluster
Ambari provides central management for starting, stopping, and reconfiguring Hadoop services across the entire cluster.
Monitor a Hadoop Cluster
Ambari provides a dashboard for monitoring health and status of the Hadoop cluster.
Ambari leverages Ambari Metrics System for metrics collection.
Ambari leverages Ambari Alert Framework for system alerting and will notify you when your attention is needed (e.g., a node goes down, remaining disk space is low, etc).

1) What is Apache Ambari?

Answer)Apache Ambari is an open-source software to install, manage and monitor Apache Hadoop family of components. It automates many of the basic actions performed and provides a simple and easy to use UI.

2) How does Ambari work?

Answer) Hadoop and its ecosystem of software are typically installed as a multi-node deployment. Ambari has a two level architecture of an Ambari Server and an Ambari agent. Ambari Server centrally manages all the agents and sends out operations to be performed on individual agents. Agents are installed by the server on each node (host) which in turn installs, configures and manages services in the agent

3) What are Services?

Answer) Services are the various components of the Hadoop ecosystem such as HDFS, YARN, Hive, HBase, Oozie, Druid, etc. One of the most popular open-source Hadoop distributions is the Hortonworks Data Platform (HDP)

4) How is a stack like HDP installed by Ambari?

www.smartdatacamp.com

Answer) Each version of HDP corresponds to a version of Ambari which supports the HDP version.
The latest Ambari version can be ascertained from docs.hortonworks.com
Once the Ambari repository is downloaded and installed, Ambari shows the list of HDP versions it supports.
Ambari also guides the users through an installation wizard which requests the users for details like the services to be installed, on which node, etc.

5) Ok, Ambari installed HDP. What else can it do?

Answer) Ambari can also monitor and manage various services on Hadoop. For example, Ambari can start/stop services it manages, a user can add additional services, delete services, etc.
The user can also get metrics/data about the health of the various services managed by Ambari
Ambari also provides Views into some of the components like Hive, HBase, Pig, HDFS, etc., where a user can run queries and various jobs.
Ambari also provides the users to edit their the service configurations and version those configurations so that at a later point in time, they can be restored if the changed configuration causes issues.

6) Where do I download the latest repositories for Ambari?

Answer)For obtaining Ambari package with HDP cluster definitions, go to https://docs.hortonworks.com/ - select version - Apache Ambari Installation - Obtaining Public Repositories - Ambari Repositories Get the appropriate repository for the OS required

7) Can Ambari upgrade HDP? How do I decide when to upgrade? Can I upgrade only specific service?

Answer) Yes Ambari can upgrade HDP. You can upgrade when a new release of HDP is announced by Hortonworks or if you're looking for a specific feature which has landed in a new version of HDP. Upgrading only 1 service as part of cluster upgrade is not supported, however you can apply patch or maintenance upgrades to 2.6.4.x stack to a specific service.

8)Does Ambari support other stacks like HDF?

Answer)Yes. Other than HDP, Ambari paackage from Hortonworks supports other stacks like HCP.

9)How do I secure my cluster using Ambari?

Answer)Kerberos authentication can be enabled from Ambari for network security
Install Ranger and Configure basic authorization in Ranger from Ambari
Ambari can be configured to use Knox SSO
You can setup SSL for Ambari

10)Does Ambari support HA?

Answer) Not as of now. However, one can setup an active-passive ambari-server instance. Refer to the article for more details. Ambari Server HA is planned in a future release of Ambari: AMBARI-17126

11)Where is the Ambari codebase? I heard its open source

Answer)Apache Ambari is completely open source with an Apache license. The code base is available in github.

12) How can I contribute to Ambari?

Answer) This wiki document explains how to contribute to Ambari

13) I want to perform scheduled maintenance on some of my cluster nodes? How will Ambari react to it? Stuff like adding a disk, replacing a node etc.

Answer)In Ambari, there is a maintenance mode option for all the services/hosts managed by it. One can switch on maintenance mode for the host/service affected by the maintenance which suppresses the alerts, and safely perform the maintenance operations.

14) How does Ambari decide the order in which various components should be installed on respective nodes?

Answer)Within Ambari, there is a finite state machine and a command orchestrator which manages all the dependencies of various components within it.

15)What is the significance of “ambari-qa” user?

www.smartdatacamp.com

Answer)'ambari-qa' user account is created by Ambari on all nodes in the cluster. This user performs a service check against cluster services as part of the install process. You can refer to the list of other users created while cluster installation.

16)I changed a config in a service and Ambari provided some recommendations for changes in other services, where are such recommendations coming from?

Answer)These recommendations are provided by a component called StackAdvisor. It is responsible for recommending various configurations at installation time and also maintaining the dependencies for the various services managed by Ambari.

17)How do I customize the configurations in Ambari Server?

Answer) ambari.properties is located at /etc/conf/ambari-server/ambari.properties
There are a set of properties with jdbc in the key. This is to configure the ambari database.
There are another set of properties related to jdk and configuring the java version for ambari
Another set of properties starting with “views” for configuring behaviour of ambari views.
Security related configurations appear with the keyword “kerberos”, “security”, “jce”, etc
You can run the ambari-server as a non-root user by specifying the username in “ambari-server.user”
You can also specify timeouts for the common ambari installation tasks, e.g.: agent.package.install.task.timeout, agent.service.check.task.timeout, agent.task.timeout, server.task.timeout
One can also set the time an Ambari login can be active by specifying the time in server.http.session.inactive_timeout

18)Can Ambari manage more than one cluster?

Answer)As of now, an Ambari instance can manage only one cluster. However, you can remotely view the “views” of another cluster in the same instance. You can read this blog post for more information

19)I have a Hadoop cluster. How can I start managing under Ambari ?

Answer)If the cluster is not yet in production, clean up the cluster and install the cluster from scratch using Ambari, (after backing up the data, of course).
If it production critical, then:
Setup ambari-server and ambari database
Install Update ambari-agents to point to the ambari-server
Use Ambari APIs to perform cluster takeover i.e. add cluster, add hosts, register services and components, register host components. Refer here for Ambari APIs

An alternative is to create an Ambari blueprint based on the current configuration and install the Cluster on Ambari using the blueprint.

20)Does Ambari authentication work with SSO?

Answer)Yes. You can use Knox SSO for connecting to an IDP for Ambari authentication.

21)What is first place to start troubleshooting an Ambari issue?

Answer)Verify if ambari-server is up and running and ambari-server is able to communicate to all the ambari-agents.
Perform a ambari database consistency check to make sure there are no database consistency errors. Run the following command on the ambari-server: ambari-server check-database
Ambari server logs available at /var/log/ambari-server/ambari-server.log
Ambari agent logs available at /var/log/ambari-agent/ambari-agent.log
Ambari Agent task logs on any host with an Ambari Agent: /var/lib/ambari-agent/data/
This location contains logs for all tasks executed on an Ambari Agent host. Each log name includes:
command-N.json - the command file corresponding to a specific task.
output-N.txt - the output from the command execution.
errors-N.txt - error messages.
You can configure the logging level for ambari-server.log by modifying /etc/ambari-server/conf/log4j.properties on the Ambari Server host. For the Ambari Agents, you can set the loglevel in/etc/ambari-agent/conf/ambari-agent.ini on each host running an Ambari Agent.
You could also take a look at the troubleshooting guide for specific issues while installation, usage/upgrading a cluster using Ambari

22)HDP installation via Ambari failed. What options do I have?

Answer)Try to re-run the steps from the Ambari console
Restore to a previous snapshot, if available
If your issue is not yet resolved, raise a support case if you're a Hortonworks customer or post a question on HCC for further help

23)What if Ambari server host crashes? Recovery options?

Answer)Maintaining a backup of Ambari Database for any changes to the cluster configuration is always recommended.
If a backup is maintained, you can recover the host and install ambari-server afresh by pointing to the recovered database.

www.smartdatacamp.com

If there is no backup, Ambari takeover can be performed by manually adding the hosts, cluster and services installed via Ambari APIs. Refer here for list of Ambari APIs and their functions

24)What happens when a node in a cluster running a master service component crashes?

Answer)One can attempt to recover the host via the 'Recover Host' option from the Ambari Web UI.

25)What happens when a node in a cluster running a slave service component crashes?

Answer)One can attempt to recover the node (after recovering it manually) by performing the action 'Recover Host' from the Ambari UI.
If the above action does not restore the cluster to its original state, follow the following steps:
Clean up the ambari-agent and all other files on the node.
Perform the 'Add Host' operation via Ambari UI to register the node as a new Node
Select the master/slave components to be installed as part of the 'Add Host' wizard

Apache Apex

Apache Apex is a Hadoop YARN native big data processing platform, enabling real time stream as well as batch processing for your big data.

1) What is the differences between Apache Spark and Apache Apex?

Answer)Apache Spark is actually a batch processing. If you consider Spark streaming (which uses spark underneath) then it is micro-batch processing. In contrast, Apache apex is a true stream processing. In a sense that, incoming record does NOT have to wait for next record for processing. Record is processed and sent to next level of processing as soon as it arrives.

2)How Apache Apex is different from Apache Storm?

Answer) There are fundamental differences in architecture which make each of the platform very different in terms of latency, scaling and state management.
At the very basic level,
Apache Storm uses record acknowledgement to guarantee message delivery.
Apache Apex uses checkpointing to guarantee message delivery.

3) How to calculate network latency between operators in Apache Apex

Answer) Assuming your tuples are strings and that the clocks on your cluster nodes are synchronized, you can append a timestamp to each tuple in the sending operator. Then, in the receiving operator, you can strip out the timestamp and compare it to the current time. You can, of course, suitably adapt this approach for other types. If averaged over a suitably large number of tuples, it should give you a good approximation of the network latency.

4) Can an Input Operator be used in the middle of a DAG in Apache Apex

Answer) This is an interesting use-case. You should be able to extend an input operator (say JdbcInputOperator since you want to read from a database) and add an input port to it. This input port receives data (tuples) from another operator from your DAG and updates the "where" clause of the JdbcInputOperator so it reads the data based on that. Hope that is what you were looking for.

5) What is the operator lifecycle in Apache Apex?

www.smartdatacamp.com

Answer) A given operator has the following life cycle as below. The life cycle spans over the execution period of the instance of the operator. In case of operator failure, the lifecycle starts over as below. A checkpoint of operator state occurs periodically once every few windows and it becomes the last known checkpoint in case of failure.

→ Constructor is called
→ State is applied from last known checkpoint
→ setup()
→ loop over {
→ beginWindow()
→ loop over {
→ process()
}
→ endWindow()
}
→ teardown()

6) How to restart Apache Apex application?

Answer)Apache Apex provides a command line interface, "apex" (previously called "dtcli") script, to interact with the applications. Once an application is shut down or killed, you can restart it using following command:
launch pi-demo-3.4.0-incubating-SNAPSHOT.apa -originalAppId application_1465560538823_0074 -Ddt.attr.APPLICATION_NAME="Relaunched PiDemo" -exactMatch "PiDemo"
where,
-originalAppId is ID of the original app. This will ensure that the operators continue from where the original app left-off.
-Ddt.attr.APPLICATION_NAME gives the new name for relaunched app
-exactMatch is used to specify the exact app name
Note that, -Ddt.attr.APPLICATION_NAME & -exactMatch are optional.

7) Does Apache Apex rely on HDFS or does it have its own file system?

Answer) Apache Apex uses checkpointing of operator state for fault tolerance. Apex uses HDFS to write these checkpoints for recovery. However, the store for checkpointing is configurable. Apex also has an implementation to checkpoint to Apache Geode. Apex also uses HDFS to upload artifacts such application package containing the application jar, its dependencies and configurations etc that are needed to launch the application.

8)How to pass arguments to application.java class in Apache Apex?

Answer)You can pass arguments as Configuration. This configuration will be passed as an argument to populateDAG() method in Application.java.

Configuration is org.apache.hadoop.conf.Configuration. You can specify it as xml. For xml syntax please refer to
https://hadoop.apache.org/docs/r2.6.1/api/org/apache/hadoop/conf/Configuration.html.

There are different ways in which properties can be specified:

~/.dt/dt-site.xml: By default apex cli will look for this file (~ is your home directory). You should use this file for the properties which are common to all the applications in your environment.

-conf option on apex cli: launch command on apex cli provides -conf option to specify properties. You need to specify the path for the configuration xml. You should use this file for the properties which are specific to a particular application or specific to this launch of the application.

-Dproperty-name=value: launch command on apex cli provides -D option to specify properties. You can specify multiple properties like -Dproperty-name1=value1 -Dproperty-name2=value2 etc.

9)How does Apache Apex handle back pressure?

Answer)Buffer server is a pub-sub mechanism within Apex platform that is used to stream data between operators. The buffer server always lives in the same container as the upstream operator (one buffer server per container irrespective of number of operators in container); and the output of upstream operator is written to buffer server. The current operator subscribes from the upstream operator's buffer server when a stream is connected.

So if an operator fails, the upstream operator's buffer server will have the required data state until a common checkpoint is reached. If the upstream operator fails, its upstream operator's buffer server has the data state and so on. Finally, if the input operator fails, which has no upstream buffer server, then the input operator is responsible to replay the data state. Depending on the external system, input operator either relies on the external system for replays or maintain the data state itself until a common checkpoint is reached.

If for some reason the buffer server fails, the container hosting the buffer server fails. So, all the operators in the container and their downstream operators are redeployed from last known checkpoint.

www.smartdatacamp.com

Apache Avro

Apache Avro is a data serialization system.
Avro provides:
Rich data structures.
A compact, fast, binary data format.
A container file, to store persistent data.
Remote procedure call (RPC).
Simple integration with dynamic languages. Code generation is not required to read or write data files nor to use or implement RPC protocols. Code generation as an optional optimization, only worth implementing for statically typed languages.

1) Deserializing from a byte array

Answer) This example takes a byte array containing the Avro serialization of a user and returns a User object.SpecificDatumReader<User> reader = new SpecificDatumReader<User>(User.getClassSchema());
Decoder decoder = DecoderFactory.get().binaryDecoder(bytes, null);
User user = reader.read(null, decoder);

2)Serializing to a byte array?

Answer) This example takes a User object and returns a newly allocated byte array with the Avro serialization of that user.
ByteArrayOutputStream out = new ByteArrayOutputStream();
BinaryEncoder encoder = EncoderFactory.get().binaryEncoder(out, null);
DatumWriter<User> writer = new SpecificDatumWriter<User>(User.getClassSchema());
writer.write(user, encoder);
encoder.flush();
out.close();
byte[] serializedBytes = out.toByteArray();

3)How can I serialize directly to/from a byte array?

Answer) As pointed out in the specification, Avro data should always be stored with its schema. The Avro provided classes DataFileWriter, DataFileReader, and DataFileStream all ensure this by serializing the Schema in a container header. In some special cases, such as when implementing a new storage system or writing unit tests, you may need to write and read directly with the bare Avro serialized values.

4)Why isn't every value in Avro nullable?

Answer)When serialized, if any value may be null then it must be noted that it is non-null, adding at least a bit to the size of every value stored and corresponding computational costs to create this bit on write and interpret it on read. These costs are wasted when values may not in fact be null, as is the case in many datasets. In Avro such costs are only paid when values may actually be null.
Also, allowing values to be null is a well-known source of errors. In Avro, a value declared as non-null will always be non-null and programs need not test for null values when processing it nor will they ever fail for lack of such tests.
Tony Hoare calls his invention of null references his "Billion Dollar Mistake".
http://qconlondon.com/london-2009/presentation/Null+References:+The+Billion+Dollar+Mistake
Also note that in some programming languages not all values are permitted to be null. For example, in Java, values of type boolean, byte, short, char, int, float, long, and double may not be null.

5)What is the purpose of the sync marker in the object file format?

Answer From Doug Cutting:

HDFS splits files into blocks, and mapreduce runs a map task for each block. When the task starts, it needs to be able to seek into the file to the start of the block process through the block's end. If the file were, e.g., a gzip file, this would not be possible, since gzip files must be decompressed from the start. One cannot seek into the middle of a gzip file and start decompressing. So Hadoop's SequenceFile places a marker periodically (~64k) in the file at record and compression boundaries, where processing can be sensibly started. Then, when a map task starts processing an HDFS block, it finds the first marker after the block's start and continues through the first marker in the next block of the file. This requires a bit of non-local access (~0.1%). Avro's data file uses the same method as SequenceFile.

6)More generally, how do Avro types map to Java types?

Answer)The mappings are documented in the package javadoc for generic, specific and reflect API.

7)How are Strings represented in Java?

Answer)They use org.apache.avro.util.Utf8, not java.lang.String.

8)How do I statically compile a schema or protocol into generated code?

www.smartdatacamp.com

Answer) In Java
Add the avro jar, the jackson-mapper-asl.jar and jackson-core-asl.jar to your CLASSPATH.
Run java org.apache.avro.specific.SpecificCompiler <json file>.
This appears to be out of date, the SpecificCompiler requires two arguments, presumably an input and and output file, but it isn't clear that this does.
Or use the Schema or Protocol Ant tasks. Avro's build.xml provides examples of how these are used.
Lastly, you can also use the "avro-tools" jar which ships with an Avro release. Just use the "compile (schema|protocol)" command.

9)What is Avro?

Answer)Avro is a data serialization system.

Avro provides:
Rich data structures.
A compact, fast, binary data format.
A container file, to store persistent data.
Remote procedure call (RPC).
Simple integration with dynamic languages. Code generation is not required to read or write data files nor to use or implement RPC protocols. Code generation as an optional optimization, only worth implementing for statically typed languages.

Apache Beam

Apache Beam is an open source, unified model for defining both batch and streaming data-parallel processing pipelines. Using one of the open source Beam SDKs, you build a program that defines the pipeline. The pipeline is then executed by one of Beam's supported distributed processing back-ends, which include Apache Apex, Apache Flink, Apache Spark, and Google Cloud Dataflow.

1) What are the benefits of Apache Beam over Spark/Flink for batch processing?

Answer)There's a few things that Beam adds over many of the existing engines.

Unifying batch and streaming. Many systems can handle both batch and streaming, but they often do so via separate APIs. But in Beam, batch and streaming are just two points on a spectrum of latency, completeness, and cost. There's no learning/rewriting cliff from batch to streaming. So if you write a batch pipeline today but tomorrow your latency needs change, it's incredibly easy to adjust. You can see this kind of journey in the Mobile Gaming examples.

APIs that raise the level of abstraction: Beam's APIs focus on capturing properties of your data and your logic, instead of letting details of the underlying runtime leak through. This is both key for portability (see next paragraph) and can also give runtimes a lot of flexibility in how they execute. Something like ParDo fusion (aka function composition) is a pretty basic optimization that the vast majority of runners already do. Other optimizations are still being implemented for some runners. For example, Beam's Source APIs are specifically built to avoid overspecification the sharding within a pipeline. Instead, they give runners the right hooks to dynamically rebalance work across available machines. This can make a huge difference in performance by essentially eliminating straggler shards. In general, the more smarts we can build into the runners, the better off we'll be. Even the most careful hand tuning will fail as data, code, and environments shift.

Portability across runtimes.: Because data shapes and runtime requirements are neatly separated, the same pipeline can be run in multiple ways. And that means that you don't end up rewriting code when you have to move from on-prem to the cloud or from a tried and true system to something on the cutting edge. You can very easily compare options to find the mix of environment and performance that works best for your current needs. And that might be a mix of things -- processing sensitive data on premise with an open source runner and processing other data on a managed service in the cloud.

Designing the Beam model to be a useful abstraction over many, different engines is tricky. Beam is neither the intersection of the functionality of all the engines (too limited!) nor the union (too much of a kitchen sink!). Instead, Beam tries to be at the forefront of where data processing is going, both pushing functionality into and pulling patterns out of the runtime engines.

Keyed State is a great example of functionality that existed in various engines and enabled interesting and common use cases, but wasn't originally expressible in Beam. We recently expanded the Beam model to include a version of this functionality according to Beam's design principles.

www.smartdatacamp.com

And vice versa, we hope that Beam will influence the roadmaps of various engines as well. For example, the semantics of Flink's DataStreams were influenced by the Beam (née Dataflow) model.

This also means that the capabilities will not always be exactly the same across different Beam runners at a given point in time. So that's why we're using capability matrix to try to clearly communicate the state of things.

2)Apache Beam : FlatMap vs Map?

Answer)These transforms in Beam are exactly same as Spark (Scala too).

A Map transform, maps from a PCollection of N elements into another PCollection of N elements.

A FlatMap transform maps a PCollections of N elements into N collections of zero or more elements, which are then flattened into a single PCollection.

As a simple example, the following happens:

beam.Create([1, 2, 3]) | beam.Map(lambda x: [x, 'any'])

The result is a collection of THREE lists: [[1, 'any'], [2, 'any'], [3, 'any']]

Whereas:

beam.Create([1, 2, 3]) | beam.FlatMap(lambda x: [x, 'any'])

The lists that are output by the lambda, are then flattened into a

collection of SIX single elements: [1, 'any', 2, 'any', 3, 'any']

3)How do you express denormalization joins in Apache Beam that stretch over long periods of time

Answer)Since Producer may appear years before its Product, you can use some external storage (e.g. BigTable) to store your Producers and write a ParDo for Product stream to do lookups and perform join. To further optimize performance, you can take advantage of stateful DoFn feature to batch lookups.

You can still use windowing and CoGroupByKey to do join for cases where Product data is delivered before Producer data. However, the window here can be small enough just to handle out-of-order delivery.

4)Apache Airflow or Apache Beam for data processing and job scheduling

Answer)Airflow can do anything. It has BashOperator and PythonOperator which means it can run any bash script or any Python script.
It is a way to organize (setup complicated data pipeline DAGs), schedule, monitor, trigger re-runs of data pipelines, in a easy-to-view and use UI.

Also, it is easy to setup and everything is in familiar Python code.
Doing pipelines in an organized manner (i.e using Airflow) means you don't waste time debugging a mess of data processing (cron) scripts all over the place.
Apache Beam is a wrapper for the many data processing frameworks (Spark, Flink etc.) out there. The intent is so you just learn Beam and can run on multiple backends (Beam runners).
If you are familiar with Keras and TensorFlow/Theano/Torch, the relationship between Keras and its backends is similar to the relationship between Beam and its data processing backends.

5)What are the use cases for Apache Beam and Apache Nifi? It seems both of them are data flow engines. In case both have similar use case, which of the two is better?

Answer)Apache Beam is an abstraction layer for stream processing systems like Apache Flink, Apache Spark (streaming), Apache Apex, and Apache Storm. It lets you write your code against a standard API, and then execute the code using any of the underlying platforms. So theoretically, if you wrote your code against the Beam API, that code could run on Flink or Spark Streaming without any code changes.

Apache NiFi is a data flow tool that is focused on moving data between systems, all the way from very small edge devices with the use of MiNiFi, back to the larger data centers with NiFi. NiFi's focus is on capabilities like visual command and control, filtering of data, enrichment of data, data provenance, and security, just to name a few. With NiFi, you aren't writing code and deploying it as a job, you are building a living data flow through the UI that is taking effect with each action.

Stream- processing platforms are often focused on computations involving joins of streams and windowing operations. Where as a data flow tool is often complimentary and used to manage the flow of data from the sources to the processing platforms.

There are actually several integration points between NiFi and stream processing systems... there are components for Flink, Spark, Storm, and Apex that can pull data from NiFi, or push data back to NiFi. Another common pattern would be to use MiNiFi + NiFi to get data into Apache Kafka, and then have the stream processing systems consume from Kafka.

www.smartdatacamp.com

Bigtop

Bigtop is a project for the development of packaging and tests of the Apache Hadoop ecosystem.

1) What is BigTop?

Answer)Bigtop is a project for the development of packaging and tests of the Apache Hadoop ecosystem.

The primary goal of Bigtop is to build a community around the packaging and interoperability testing of Hadoop-related projects. This includes testing at various levels (packaging, platform, runtime, upgrade, etc...) developed by a community with a focus on the system as a whole, rather than individual projects.

Build, packaging and integration test code that depends upon official releases of the Apache Hadoop-related projects (HDFS, MapReduce, HBase, Hive, Pig, ZooKeeper, etc...) will be developed and released by this project. As bugs and other issues are found we expect these to be fixed upstream

2)Does BigTop patches source releases?

Answer)BigTop does NOT patch any source release and does NOT have any mechanism to deal with anything else than bare source and pristine releases. NO patches will be applied. NOT even for build or security issues.

3)How to add a component to BigTop?

Answer)First, you need to add an entry for your project in bigtop.mk similar to what is there for the others

Put all the files common to RPM or DEB packaging in src/pkg/common/<YOUR_PROJECT_NAME>/. This may include, but not limited to, common build or installation scripts or service scripts for daemons

Put your spec file in src/pkg/rpm/<YOUR_PROJECT_NAME>/SPECS/

Put any additional file needed for the creation of your project's RPMs in src/pkg/rpm/<YOUR_PROJECT_NAME>/SOURCES/

Put all your files needed for the creation of your project's DEBs in src/pkg/deb/<YOUR_PROJECT_NAME>/SPECS/

4)How to build a component of BigTop?

Answer)You must type from a command line:

make <YOUR_PROJECT_NAME>-<TARGET>

Where <TARGET> can be:

sdeb if you wish to build source DEBs

deb if you wish to build DEBs

srpm if you wish to build source RPMs

rpm if you wish to build RPMs

apt if you wish to build a repository for the already built DEBs

yum if you wish to build a repository for the already built RPMs. Note this creates a repomd repository which will only work for GNU/Linux distributions of the Fedora/CentOS/RHEL/openSUSE family

www.smartdatacamp.com

Apache Calcite

Apache Calcite is a dynamic data management framework.

It contains many of the pieces that comprise a typical database management system, but omits some key functions: storage of data, algorithms to process data, and a repository for storing metadata.

Calcite intentionally stays out of the business of storing and processing data. As we shall see, this makes it an excellent choice for mediating between applications and one or more data storage locations and data processing engines. It is also a perfect foundation for building a database: just add data.

1)How to push down project, filter, aggregation to TableScan in Calcite

Answer)Creating a new RelOptRule is the way to go. Note that you shouldn't be trying directly remove any nodes inside a rule. Instead, you match a subtree that contains the nodes you want to replace (for example, a Filter on top of a TableScan). And then replace that entire subtree with an equivalent node which pushes down the filter.

This is normally handled by creating a subclass of the relevant operation which conforms to the calling convention of the particular adapter. For example, in the Cassandra adapter, there is a CassandraFilterRule which matches a LogicalFilter on top of a CassandraTableScan. The convert function then constructs a CassandraFilter instance. The CassandraFilter instance sets up the necessary information so that when the query is actually issued, the filter is available.

Browsing some of the code for the Cassandra, MongoDB, or Elasticsearch adapters may be helpful as they are on the simpler side. I would also suggest bringing this to the mailing list as you'll probably get more detailed advice there.

2)I would like to use the apache calcite api raw without using jdbc connections. I can use the jdbc api just fine but I am getting null pointer exceptions when trying to use the api.

Answer)There's some crazy stuff going on here apparently. You need to pass internalParameters that you get out of the prepare call into your DataContext, and look them up in get. Apparently Calcite uses this to pass the query object around. You probably want to implement the other DataContext keys (current time, etc) as well.

```
final class MyDataContext(rootSchema: SchemaPlus, map: util.Map[String, Object])
extends DataContext {
override def get(name: String): AnyRef = map.get(name)

...
}
```

```
// ctx is your AdapterContext from above
val prepared = new CalcitePrepareImpl().prepareSql(ctx, query, classOf[Array[Object]], -1)
val dataContext = new DerpDataContext(
ctx.getRootSchema.plus(),
prepared.internalParameters
)
```

3)How to change Calcite's default sql grammar, to support such sql statement "select func(id) as (a, b, c) from xx;"

Answer)To change the grammar accepted by the SQL parser, you will need to change the parser. There are two ways of doing this.

The first is to fork the project and change the core grammar, Parser.jj. But as always when you fork a project, you are responsible for re-applying your changes each time you upgrade to a new version of the project.

The second is to use one of the grammar expansion points provided by the Calcite project. Calcite's grammar is written in JavaCC, but the it first runs the grammar though the FreeMarker template engine. The expansion points are variables in the template that your project can re-assign. For example, if you want to add a new DDL command, you can modify the createStatementParserMethods variable, as is done in Calcite's parser extension test:

List of methods for parsing extensions to "CREATE [OR REPLACE]" calls.

Each must accept arguments "(Span span, boolean replace)".

createStatementParserMethods: [

"SqlCreateTable"

]

Which of these approaches to use? Definitely use the second if you can, that is, if your grammar change occurs in one of the pre-defined expansion points. Use the first if only if you must, because you will run into the problem of maintaining a fork of the grammar.

If possible, see whether Calcite will accept the changes as a contribution. This is the ideal scenario for you, because Calcite will take on responsibility for maintaining your grammar extension. But they probably will only accept your change if it is standard SQL or a useful feature implemented by one or more major databases. And they will require your code to be high quality and accompanied by tests.

4)I have a simple application that does text substitution on literals in the WHERE clause of a SELECT statement. I run SqlParser.parseQuery() and apply .getWhere() to the result. However, for the following query the root node is not an SqlSelect, but an SqlOrderBy:

```
select EventID, Subject
from WorkOrder
where OwnerID = 100 and Active = 1 and Type = 2
```

order by Subject
If we use "group by" instead of "order by" then the root is an SqlSelect as expected.

Is this the intended behaviour?

Answer)Yes, this is intended. ORDER BY is not really a clause of SELECT. Consider

SELECT deptno FROM Emp
UNION
SELECT deptno FROM Dept
ORDER BY 1

The ORDER BY clause applies to the whole UNION, not to the second SELECT. Therefore we made it a standalone node.

When you ask Calcite to parse a query, the top-level nodes returned can be a SqlSelect (SELECT), SqlOrderBy (ORDER BY), SqlBasicCall (UNION, INTERSECT, EXCEPT or VALUES) or SqlWith (WITH).

Apache Camel

Apache Camel is a powerful open source integration framework based on known Enterprise Integration Patterns with powerful bean integration.

1)Does Camel work on IBMs JDK

Answer)Yes, It has been tested Camel with IBM's JDK on the AIX and Linux platforms. There are a few things to look out for though.
EXCEPTION USING CAMEL-HTTP
BUILDING CAMEL-SPRING COMPONENT
RUBY SCRIPTING SUPPORT

2)How does Camel compare to Mule?

Answer)The main differences are as follows:

Camel uses a Java Domain Specific Language in addition to Spring XML for configuring the routing rules and providing Enterprise Integration Patterns

Camel's API is smaller & cleaner (IMHO) and is closely aligned with the APIs of JBI, CXF and JMS; based around message exchanges (with in and optional out messages) which more closely maps to REST, WS, WSDL & JBI than the UMO model Mule is based on

Camel allows the underlying transport details to be easily exposed (e.g. the JmsExchange, JbiExchange, HttpExchange objects expose all the underlying transport information & behaviour if its required). See How does the Camel API compare to

Camel supports an implicit Type Converter in the core API to make it simpler to connect components together requiring different types of payload & headers

Camel uses the Apache 2 License rather than Mule's more restrictive commercial license

3)How does Camel compare to servicemix?

Answer)Camel is smart routing and mediation engine which implements the Enterprise Integration Patterns and is designed to be used either inside an ESB like ServiceMix, in a Message Broker like ActiveMQ or in a smart endpoint or web services framework like CXF. ServiceMix is an ESB, a JBI container and an integration platform. So they both address different needs though they are both designed to work great together.

Camel can be deployed as a component within ServiceMix to provide EIP routing and mediation between existing JBI components together with communicating with any of the other Camel.

www.smartdatacamp.com

Components along with defining new JBI components on the NMR. So Camel is similar to the ServiceMix EIP component.

To work with Camel and ServiceMix you take your Camel Spring configuration and turn it into a JBI Service Unit using the maven plugin or archetype. For more details see ServiceMix Camel plugin.

So you could start out using Camel routing inside your application via Java or Spring; then later on if you choose to you could wrap up your routing and mediation rules as a JBI deployment unit and drop it into your ServiceMix ESB. This provides a nice agile approach to integration; start small & simple on an endpoint then as and when you need to migrate your integration components into your ESB for more centralised management, governance and operational monitoring etc.

4)How does Camel compare to ServiceMix EIP?

Answer)ServiceMix EIP was the ancestor though they both do similar things.

The main difference with ServiceMix EIP is its integrated into the existing ServiceMix XBean XML configuration whereas Camel has more Enterprise Integration Patterns and can be used outside of JBI (e.g. just with pure JMS or MINA). Also Camel supports a Java DSL or XML configuration.

CONVERTING FROM SERVICEMIX EIP TO CAMEL

5)How does Camel compare to Synapse?

We are Camel developers so take what you read here with a pinch of salt. If you want to read a less biased comparison try reading this review which has a slight Synapse bias since the author mostly uses Synapse :smile:

However we think the main differences are:

the Camel community is way more active according to the nabble statistics (Synapse is inside the Apache Web Services bar) or by comparing Camel and Synapse on markmail.

Camel is the default routing engine included in Apache ActiveMQ for Message Orientated middleware with EIP and Apache ServiceMix the ESB based around OSGi and JBI at Apache - both of which are very popular too.

Camel is designed from the ground up around Enterprise Integration Patterns — having an EIP pattern language implemented in Java and Spring XML.

Camel is designed to work with pretty much all kinds of transport as well as working with any Data Format. When we first looked at Synapse it was based around Axis 2 and WS-* though apparently thats no longer the case.

6)What is Camel

Answer)Apache Camel is a versatile open-source integration framework based on known Enterprise Integration Patterns.

Camel empowers you to define routing and mediation rules in a variety of domain-specific languages, including a Java-based Fluent API, Spring or Blueprint XML Configuration files. This means you get smart completion of routing rules in your IDE, whether in a Java or XML editor.

Apache Camel uses URIs to work directly with any kind of Transport or messaging model such as HTTP, ActiveMQ, JMS, JBI, SCA, MINA or CXF, as well as pluggable Components and Data Format options. Apache Camel is a small library with minimal dependencies for easy embedding in any Java application. Apache Camel lets you work with the same API regardless which kind of Transport is used — so learn the API once and you can interact with all the Components provided out-of-box.

Apache Camel provides support for Bean Binding and seamless integration with popular frameworks such as CDI, Spring and Blueprint. Camel also has extensive support for unit testing your routes.

The following projects can leverage Apache Camel as a routing and mediation engine:
Apache ServiceMix — a popular distributed open source ESB and JBI container
Apache ActiveMQ — a mature, widely used open source message broker
Apache CXF — a smart web services suite (JAX-WS and JAX-RS)
Apache Karaf — a small OSGi based runtime in which applications can be deployed
Apache MINA — a high-performance NIO-driven networking framework

7)How do I specify which method to use when using beans in routes?

Answer)However if you have overloaded methods you need to specify which of those overloaded method you want to use by specifying parameter type qualifiers.

8)How can I get the remote connection IP address from the camel-cxf consumer ?

Answer)From Camel 2.6.0, you can access the CXF Message by using the key of CamelCxfMessage from message header, and you can get the ServletRequest instance from the CXF message, then you can get the remote connection IP easily.

Here is the code snippet:

```
// check the remote IP from the CXF Message
org.apache.cxf.message.Message cxfMessage =
exchange.getIn().getHeader(CxfConstants.CAMEL_CXF_MESSAGE,
org.apache.cxf.message.Message.class);
ServletRequest request = (ServletRequest) cxfMessage.get("HTTP.REQUEST");
String remoteAddress = request.getRemoteAddr();
```

www.smartdatacamp.com

8)How can I stop a route from a route?

Answer)The CamelContext provides API for managing routes at runtime. It has a stopRoute(id) and startRoute(id) methods.

Stopping a route during routing an existing message is a bit tricky. The reason for that is Camel will Graceful Shutdown the route you are stopping. And if you do that while a message is being routed the Graceful Shutdown will try to wait until that message has been processed.

The best practice for stopping a route from a route, is to either:

signal to another thread to stop the route

spin off a new thread to stop the route

Using another thread to stop the route is also what is normally used when stopping Camel itself, or for example when an application in a server is stopped etc. Its too tricky and hard to stop a route using the same thread that currently is processing a message from the route. This is not advised to do, and can cause unforeseen side effects.

9)How does Camel work?

Answer)Camel uses a Java based Routing Domain Specific Language (DSL) or an XML Configuration to configure routing and mediation rules which are added to a CamelContext to implement the various Enterprise Integration Patterns.

At a high level Camel consists of a CamelContext which contains a collection of Component instances. A Component is essentially a factory of Endpoint instances. You can explicitly configure Component instances in Java code or an IoC container like Spring or Guice, or they can be auto-discovered using URIs.

An Endpoint acts rather like a URI or URL in a web application or a Destination in a JMS system; you can communicate with an endpoint; either sending messages to it or consuming messages from it. You can then create a Producer or Consumer on an Endpoint to exchange messages with it.

The DSL makes heavy use of pluggable Languages to create an Expression or Predicate to make a truly powerful DSL which is extensible to the most suitable language depending on your needs. Many of the Languages are also supported as Annotation Based Expression Language.

10)How does Camel work with ActiveMQ?

Answer)You can use Camel to do smart routing and implement the Enterprise Integration Patterns inside:
the ActiveMQ message broker
the ActiveMQ JMS client

So Camel can route messages to and from Mail, File, FTP, JPA, XMPP other JMS providers and any of the other Camel Components as well as implementating all of the Enterprise Integration Patterns such as Content Based Router or Message Translator.

11)How does Camel work with ServiceMix?

Answer)You can use Camel to do smart routing and implement the Enterprise Integration Patterns inside of the JBI container, routing between existing JBI components together with communicating with any of the other Camel Components.

To do this you take your Camel Spring configuration and turn it into a JBI Service Unit using the maven plugin or archetype.

12)How can I get the remote connection IP address from the camel-cxf consumer ?

Answer)From Camel 2.6.0, you can access the CXF Message by using the key of CamelCxfMessage from message header, and you can get the ServletRequest instance from the CXF message, then you can get the remote connection IP easily.

Here is the code snippet:

```
// check the remote IP from the CXF Message
org.apache.cxf.message.Message cxfMessage =
exchange.getIn().getHeader(CxfConstants.CAMEL_CXF_MESSAGE,
org.apache.cxf.message.Message.class);
ServletRequest request = (ServletRequest) cxfMessage.get("HTTP.REQUEST");
String remoteAddress = request.getRemoteAddr();
```

13)How does Camel look up beans and endpoints?

Answer)There are many times using Camel that a name is used for a bean such as using the Bean endpoint or using the Bean Language to create a Expression or Predicate or referring to any Component or Endpoint.

Camel uses the Registry to resolve names when looking up beans or components or endpoints. Typically this will be Spring; though you can use Camel without Spring in which case it will use the JNDI registry implementation.

Lots of test cases in the camel-core module don't use Spring (as camel-core explicitly doesn't depend on spring) - though test cases in camel-spring do.

So you can just define beans, components or endpoints in your Registry implementation then you can refer to them by name in the Endpoint URIs or Bean endpoints or Bean Language expressions.

www.smartdatacamp.com

13)How do I change the logging?

Answer)We use commons-logging to log information in the broker client and the broker itself so you can fully configure the logging levels desired, whether to log to files or the console, as well as the underlying logging implementation (Log4J, Java SE logger, etc.) you wish to use. For Log4J, full instructions are in its manual, but in a nutshell:

Add log4j.jar to your classpath

Create a log4j.properties file specifying desired logging configuration (The Camel distribution has example log4j.properties files you can use — see for example in the /examples/camel-example-as2/src/main/resources folder.)

Place the log4j.properties file in the folder where the compiled .class files are located (typically the classes folder) — this will place the properties file on the classpath, where it needs to be at runtime.

14)How do I configure endpoints?

Answer)There are a few different approaches to configuring components and endpoints.

USING JAVA CODE

You can explicitly configure a Component using Java code as shown in this example

Or you can explicitly get hold of an Endpoint and configure it using Java code as shown in the Mock endpoint examples.

SomeEndpoint endpoint = camelContext.getEndpoint("someURI", SomeEndpoint.class);

endpoint.setSomething("aValue");

14)How do I configure password options on Camel endpoints without the value being encoded?

Answer)When you configure Camel endpoints using URIs then the parameter values gets url encoded by default.

This can be a problem when you want to configure passwords as is.

To do that you can tell Camel to use the raw value, by enclosing the value with RAW(value).

15)How do I configure the default maximum cache size for ProducerCache or ProducerTemplate?

Answer)This applies to ConsumerCache and ConsumerTemplate as well.

You can configure the default maximum cache size by setting the Exchange.MAXIMUM_CACHE_POOL_SIZE property on CamelContext.
getCamelContext().getProperties().put(Exchange.MAXIMUM_CACHE_POOL_SIZE, "50");
And in Spring XML its done as:
<camelContext>
<properties>
<property key="CamelMaximumCachePoolSize" value="50"/>
</properties>
...
</camelContext>

The default maximum cache size is 1000.

At runtime you can see the ProducerCache in JMX as they are listed in the services category.

15)How do I configure the maximum endpoint cache size for CamelContext?

Answer)CamelContext will by default cache the last 1000 used endpoints (based on a LRUCache).
CONFIGURING CACHE SIZE
Available as of Camel 2.8
You can configure the default maximum cache size by setting the Exchange.MAXIMUM_ENDPOINT_CACHE_SIZE property on CamelContext.
getCamelContext().getProperties().put(Exchange.MAXIMUM_ENDPOINT_CACHE_SIZE, "500");
You need to configure this before CamelContext is started.
And in Spring XML its done as:
<camelContext>
<properties>
<property key="CamelMaximumEndpointCacheSize" value="500"/>
</properties>
...
</camelContext>
At runtime you can see the EndpointRegistry in JMX as they are listed in the services category.

16)How do I debug my route?

Answer)If you've created a route and its not doing what you think it is you could try using one of these features from version 1.4 onwards:
Tracer to trace in commons-logging / log4j each step that Camel takes
Debugger to let you set breakpoints at points in the route and examine historic message exchanges
Debug from your unit test if you use the Camel camel-test component

16)How do I handle failures when consuming for example from a FTP server?

www.smartdatacamp.com

Answer)When you do a route such as:
from("ftp://foo@somesever.com?password=secret").to("bean:logic?method=doSomething");
And there is a failure with connecting to the remote FTP server. The existing Error handling in Camel is based on when a message is being routed. In this case the error occurs before a message has been initiated and routed. So how can I control the error handling?
The FTP component have a few options (maximumReconnectAttempts, reconnectDelay to control number of retries and delay in between.
But you can also plugin your own implementation and determine what to do using the pollStrategy option which has more documentation Polling Consumer. Notice that the option pollStrategy applies for all consumers which is a ScheduledPollConsumer consumer. The page lists those.

16)How do I retry failed messages forever?

Answer)If you want to keep the bad message in the original queue, then you are also blocking the messages that has arrived on the queue after the bad message.
By default Camel will retry consuming a message up til 6 times before its moved to the default dead letter queue.
If you configure the Dead Letter Channel to use maximumRedeliveries = -1 then Camel will retry forever.
When you consume a message you can check the in message header org.apache.camel.redeliveryCount that contains the number of times it has been redlivered.
Or org.apache.camel.Redelivered that contains a boolean if its redelivered or if its the first time the message is processed.

16)How should I invoke my POJOs or Spring Services?

Answer)The various options are described in detail in Bean Integration, in particular the Bean Binding describes how we invoke a bean inside a route.
See the POJO Consuming for examples using either the @Consume annotation or using the routing DSL:
from("jms:someQueue").bean(MyBean.class, "someMethod");

17)How should I package applications using Camel and ActiveMQ?

Answer)So you may wish to use Camel's Enterprise Integration Patterns inside the ActiveMQ Broker. In which case the stand alone broker is already packaged to work with Camel out of the box; just add your EIP routing rules to ActiveMQ's XML Configuration like the example routing rule which ships with ActiveMQ 5.x or later. If you want to include some Java routing rules, then just add your jar to somewhere inside ActiveMQ's lib directory.

If you wish to use ActiveMQ and/or Camel in a standalone application, we recommend you just create a normal Spring application; then add the necessary jars and customise the Spring XML and you're good to go.

18)How to define a static Camel converter method in Scala?

Answer)When you use Scala object you can define the static method for others to use. Scala will create a class which implements the singleton pattern for that class object.

If the object name is A, you can find the singleton class name with A$. Using javap to recompile the class A and A$, you will find A has bunch of static method, and A$ doesn't have any of them. If you specify the converter class package name in META-INF/service/org/apache/camel/TypeConverter, Camel will load the class A and A$ at the same time. As the A$ construction method is not supposed to be invoked, Camel will complain that he cannot load the converter method which you are supposed to use because he can't create an instance of A$.

To avoid this kind of error, we need to specify the full class name of A in the TypeConverter to let Camel load the converter directly.

19)How to send the same message to multiple endpoints?

Answer)When you need to send the same message to multiple endpoints then you should use Multicast.

In the sample below we consume messages from the activemq queue foo and want to send the same message to both seda:foo and seda:bar. Sending the same message requires that we use Multicast. This is done by adding the multicast() before the to type:

from("activemq:queue:foo").multicast().to("seda:foo", "seda:bar");

Pipeline is default in Camel

If you have a route such as:

from("activemq:queue:foo").to("seda:foo", "seda:bar");

It is by default a pipeline in Camel (that is the opposite to Multicast). In the above example using pipes and filters then the result from seda:foo is sent to seda:bar, ie. its not the same message sent to multiple destinations, but a sent through a chain (the pipes and the filters).

20)Why does FTP component not download any files?

Answer)The FTP component has many options. So make sure you have configured it properly. Also a common issue is that you have to use either active or passive mode. So you may have to set passiveMode=true on the endpoint configuration.

www.smartdatacamp.com

21)Why does useOriginalMessage with error handler not work as expected?

Answer)If you use the useOriginalMessage option from the Camel Error Handler then it matters if you use this with EIPs such as:
Recipient List
Splitter
Multicast
Then the option shareUnitOfWork on these EIPs influence the message in use by the useOriginalMessage option.

22)Why is my message body empty?

Answer)In Camel the message body can be of any types. Some types are safely readable multiple times, and therefore do not 'suffer' from becoming 'empty'. So when you message body suddenly is empty, then that is often related to using a message type that is no re-readable; in other words, the message body can only be read once. On subsequent reads the body is now empty. This happens with types that are streaming based, such as java.util.InputStream, etc.

A number of Camel components supports and use streaming types out of the box. For example the HTTP related components, CXF, etc.

Camel offers a functionality Stream caching; that caches the stream, so it can be re-readable. By enabling this cache, the message body would no longer be empty.

23)Why use multiple CamelContext?

Answer)In general, you don't tend to want multiple camel contexts in your application, if you're running Camel as a standalone Java instance. However, if you're deploying Camel routes as OSGi bundles, or WARs in an application server, then you can end up having multiple routes being deployed, each in it's own, isolated camel context, in the same JVM. This makes sense: you want each Camel application to be deployable in isolation, in it's own Application Context, and not affected by the other Camel applications.

If you want the endpoints or producers in different camel contexts to communicate with another, there are a number of solutions. You can use the ServiceMix NMR, or you can use JMS, or you can use Camel's VM transport.

24)How do I invoke Camel routes from JBI?

Answer)When you use the JBI endpoint as follows:
from("jbi:endpoint:http://foo.bar.org/MyService/MyEndpoint")
you automatically expose the endpoint to the NMR bus where service qname is:

{http://foo.bar.org}MyService
and endpoint name is MyEndpoint.
Then if you send a message via the JBI NMR to this JBI endpoint then it will be sent to the above Camel route.
Sending works in the same way. You use:
to("jbi:endpoint:http://foo.bar.org/MyService/MyEndpoint")
to send messages to JBI endpoint deployed to the bus.
I noticed that people are used to somehow 'declaring' endpoints in SMX. In Camel it is enough to simply start a flow from a jbi endpoint and Camel will create it automatically.

24)How Do I Make My JMS Endpoint Transactional?

Answer)I have a JMS route like this:
from("activemq:Some.Queue")
.bean(MyProcessor.class);

24)How do the direct, event, seda and vm endpoints compare?

Answer)VM and SEDA endpoints are basically the same; they both offer asychronous in memory SEDA queues; they differ in visibility — endpoints are visible inside the same JVM or within the same CamelContext respectively.

Direct uses no threading; it directly invokes the consumer when sending.

Spring Event adds a listener to Spring's application events; so the consumer is invoked the same thread as Spring notifies events. Event differs in that the payload should be a Spring ApplicationEvent object whereas Direct, SEDA and VM can use any payload.

24)How do the Timer and Quartz endpoints compare?

Answer)Timer is a simple, non persistence timer using the JDK's in built timer mechanism.

Quartz uses the Quartz library which uses a database to store timer events and supports distributed timers and cron notation.

24)Why does my JMS route only consume one message at once?

Answer)The default JMS endpoint configuration defines concurrentConsumers to be 1 so only 1 message is processed concurrently at any point in time. To change this to make things more concurrent, just configure this value; either at the JMS component level or endpoint level.

E.g.

www.smartdatacamp.com

from("activemq:SomeQueue?concurrentConsumers=25").

bean(SomeCode.class);

25) Why do Camel throw so many NoClassDefFoundException on startup?

Answer)Camel uses a runtime strategy to discover features while it starts up. This is used to register components, languages, type converters, etc.

If you are using the uber .jar (the big camel.jar) with all the Camel components in a single .jar filen, the this problem can typically occur. Especially the type converters is know to cause NoClassDefFoundException in the log during startup. The reasons is that some of these type converters rely on 3rd. party .jar files.

To remedy this either add the missing .jars to the classpath, or stop using the big .jar and use the fine grained jars.

www.smartdatacamp.com

Apache CarbonData

Apache CarbonData is a new big data file format for faster interactive query using advanced columnar storage, index, compression and encoding techniques to improve computing efficiency, which helps in speeding up queries by an order of magnitude faster over PetaBytes of data.

1) What are Bad Records?

Answer)Records that fail to get loaded into the CarbonData due to data type incompatibility or are empty or have incompatible format are classified as Bad Records.

2)Where are Bad Records Stored in CarbonData?

Answer)The bad records are stored at the location set in carbon.badRecords.location in carbon.properties file. By default carbon.badRecords.location specifies the following location /opt/Carbon/Spark/badrecords.

3) How to enable Bad Record Logging?

Answer) While loading data we can specify the approach to handle Bad Records. In order to analyse the cause of the Bad Records the parameter BAD_RECORDS_LOGGER_ENABLE must be set to value TRUE. There are multiple approaches to handle Bad Records which can be specified by the parameter BAD_RECORDS_ACTION.

To pass the incorrect values of the csv rows with NULL value and load the data in CarbonData, set the following in the query :'BAD_RECORDS_ACTION'='FORCE'

To write the Bad Records without passing incorrect values with NULL in the raw csv (set in the parameter carbon.badRecords.location), set the following in the query :'BAD_RECORDS_ACTION'='REDIRECT'

4) How to ignore the Bad Records?

Answer)To ignore the Bad Records from getting stored in the raw csv, we need to set the following in the query :'BAD_RECORDS_ACTION'='IGNORE'.

5) How to specify store location while creating carbon session?

www.smartdatacamp.com

Answer)The store location specified while creating carbon session is used by the CarbonData to store the meta data like the schema, dictionary files, dictionary meta data and sort indexes.

Try creating carbonsession with storepath specified in the following manner :

val carbon = SparkSession.builder().config(sc.getConf).getOrCreateCarbonSession(<carbon_store_path>)

Example:

val carbon = SparkSession.builder().config(sc.getConf).getOrCreateCarbonSession("hdfs://localhost:9000/carbon/store")

6) What is Carbon Lock Type?

Answer)The Apache CarbonData acquires lock on the files to prevent concurrent operation from modifying the same files. The lock can be of the following types depending on the storage location, for HDFS we specify it to be of type HDFSLOCK. By default it is set to type LOCALLOCK. The property carbon.lock.type configuration specifies the type of lock to be acquired during concurrent operations on table. This property can be set with the following values :

LOCALLOCK : This Lock is created on local file system as file. This lock is useful when only one spark driver (thrift server) runs on a machine and no other CarbonData spark application is launched concurrently.

HDFSLOCK : This Lock is created on HDFS file system as file. This lock is useful when multiple CarbonData spark applications are launched and no ZooKeeper is running on cluster and the HDFS supports, file based locking.

7) How to resolve Abstract Method Error?

Answer) In order to build CarbonData project it is necessary to specify the spark profile. The spark profile sets the Spark Version. You need to specify the spark version while using Maven to build project.

8)How Carbon will behave when execute insert operation in abnormal scenarios?

Answer)Carbon support insert operation, you can refer to the syntax mentioned in DML Operations on CarbonData. First, create a source table in spark-sql and load data into this created table.

```
CREATE TABLE source_table(
id String,
name String,
city String)
```

ROW FORMAT DELIMITED FIELDS TERMINATED BY ",";
SELECT * FROM source_table;
id name city
1 jack beijing
2 erlu hangzhou
3 davi shenzhen

Scenario 1 :

Suppose, the column order in carbon table is different from source table, use script "SELECT * FROM carbon table" to query, will get the column order similar as source table, rather than in carbon table's column order as expected.

CREATE TABLE IF NOT EXISTS carbon_table(
id String,
city String,
name String)
STORED AS carbondata;
INSERT INTO TABLE carbon_table SELECT * FROM source_table;
SELECT * FROM carbon_table;
id city name
1 jack beijing
2 erlu hangzhou
3 davi shenzhen

As result shows, the second column is city in carbon table, but what inside is name, such as jack. This phenomenon is same with insert data into hive table.

If you want to insert data into corresponding column in carbon table, you have to specify the column order same in insert statement.

INSERT INTO TABLE carbon_table SELECT id, city, name FROM source_table;

Scenario 2 :

Insert operation will be failed when the number of column in carbon table is different from the column specified in select statement. The following insert operation will be failed.

INSERT INTO TABLE carbon_table SELECT id, city FROM source_table;

Scenario 3 :

When the column type in carbon table is different from the column specified in select statement. The insert operation will still success, but you may get NULL in result, because NULL will be substitute value when conversion type failed.

9)Why aggregate query is not fetching data from aggregate table?

Answer)Following are the aggregate queries that won't fetch data from aggregate table:
Scenario 1 : When SubQuery predicate is present in the query.
Example:
create table gdp21(cntry smallint, gdp double, y_year date) stored as carbondata;

create datamap ag1 on table gdp21 using 'preaggregate' as select cntry, sum(gdp) from gdp21 group by cntry;
select ctry from pop1 where ctry in (select cntry from gdp21 group by cntry);
Scenario 2 : When aggregate function along with 'in' filter.
Example:
create table gdp21(cntry smallint, gdp double, y_year date) stored as carbondata;
create datamap ag1 on table gdp21 using 'preaggregate' as select cntry, sum(gdp) from gdp21 group by cntry;
select cntry, sum(gdp) from gdp21 where cntry in (select ctry from pop1) group by cntry;
Scenario 3 : When aggregate function having 'join' with equal filter.
Example:
create table gdp21(cntry smallint, gdp double, y_year date) stored as carbondata;
create datamap ag1 on table gdp21 using 'preaggregate' as select cntry, sum(gdp) from gdp21 group by cntry;
select cntry,sum(gdp) from gdp21,pop1 where cntry=ctry group by cntry;

10)Why all executors are showing success in Spark UI even after Dataload command failed at Driver side?

Answer)Spark executor shows task as failed after the maximum number of retry attempts, but loading the data having bad records and BAD_RECORDS_ACTION (carbon.bad.records.action) is set as "FAIL" will attempt only once but will send the signal to driver as failed instead of throwing the exception to retry, as there is no point to retry if bad record found and BAD_RECORDS_ACTION is set to fail. Hence the Spark executor displays this one attempt as successful but the command has actually failed to execute. Task attempts or executor logs can be checked to observe the failure reason.

11)Why different time zone result for select query output when query SDK writer output?

Answer)SDK writer is an independent entity, hence SDK writer can generate carbondata files from a non-cluster machine that has different time zones. But at cluster when those files are read, it always takes cluster time-zone. Hence, the value of timestamp and date datatype fields are not original value. If wanted to control timezone of data while writing, then set cluster's time-zone in SDK writer by calling below API.

TimeZone.setDefault(timezoneValue)

Example:

cluster timezone is Asia/Shanghai

TimeZone.setDefault(TimeZone.getTimeZone("Asia/Shanghai"))

12)How to check LRU cache memory footprint?

Answer)To observe the LRU cache memory footprint in the logs, configure the below properties in log4j.properties file.

log4j.logger.org.apache.carbondata.core.cache.CarbonLRUCache = DEBUG

This property will enable the DEBUG log for the CarbonLRUCache and UnsafeMemoryManager which will print the information of memory consumed using which the LRU cache size can be decided. Note: Enabling the DEBUG log will degrade the query performance. Ensure carbon.max.driver.lru.cache.size is configured to observe the current cache size.

Example:

18/09/26 15:05:29 DEBUG CarbonLRUCache: main Required size for entry /home/target/store/default/stored_as_carbondata_table/Fact/Part0/Segment_0/0_1537954529044.carbonindexmerge :: 181 Current cache size :: 0

18/09/26 15:05:30 INFO CarbonLRUCache: main Removed entry from InMemory lru cache :: /home/target/store/default/stored_as_carbondata_table/Fact/Part0/Segment_0/0_1537954529044.carbonindexmerge

Note: If Removed entry from InMemory LRU cache are frequently observed in logs, you may have to increase the configured LRU size.

To observe the LRU cache from heap dump, check the heap used by CarbonLRUCache class.

13)We are Getting tablestatus.lock issues When loading data ?

Answer)Symptom

17/11/11 16:48:13 ERROR LocalFileLock: main hdfs:/localhost:9000/carbon/store/default/hdfstable/tablestatus.lock (No such file or directory)

java.io.FileNotFoundException: hdfs:/localhost:9000/carbon/store/default/hdfstable/tablestatus.lock (No such file or directory)

at java.io.FileOutputStream.open0(Native Method)

at java.io.FileOutputStream.open(FileOutputStream.java:270)

at java.io.FileOutputStream.<init>(FileOutputStream.java:213)

at java.io.FileOutputStream.<init>(FileOutputStream.java:101)

Possible Cause If you use <hdfs path> as store path when creating carbonsession, may get the errors,because the default is LOCALLOCK.

Procedure Before creating carbonsession, sets as below:

import org.apache.carbondata.core.util.CarbonProperties

import org.apache.carbondata.core.constants.CarbonCommonConstants

CarbonProperties.getInstance().addProperty(CarbonCommonConstants.LOCK_TYPE, "HDFSLOCK")

www.smartdatacamp.com

14)We are Failed to load thrift libraries ?

Answer)Symptom
Thrift throws following exception :
thrift: error while loading shared libraries:
libthriftc.so.0: cannot open shared object file: No such file or directory
Possible Cause
The complete path to the directory containing the libraries is not configured correctly.
Procedure
Follow the Apache thrift docs at https://thrift.apache.org/docs/install to install thrift correctly.

15)Failed to launch the Spark Shell

Answer)Symptom
The shell prompts the following error :
org.apache.spark.sql.CarbonContext$$anon$$apache$spark$sql$catalyst$analysis
$OverrideCatalog$_setter_orgapache$spark$sql$catalyst$analysis
$OverrideCatalog$$overrides_$e
Possible Cause
The Spark Version and the selected Spark Profile do not match.
Procedure
Ensure your spark version and selected profile for spark are correct.
Use the following command :
mvn -Pspark-2.1 -Dspark.version {yourSparkVersion} clean package
Note : Refrain from using "mvn clean package" without specifying the profile.

16)Failed to execute load query on cluster

Answer)Symptom
Load query failed with the following exception:
Dictionary file is locked for updation.
Possible Cause
The carbon.properties file is not identical in all the nodes of the cluster.
Procedure
Follow the steps to ensure the carbon.properties file is consistent across all the nodes:
Copy the carbon.properties file from the master node to all the other nodes in the cluster. For example, you can use ssh to copy this file to all the nodes.
For the changes to take effect, restart the Spark cluster.

17)Failed to execute insert query on cluster

Answer)Symptom

Load query failed with the following exception:
Dictionary file is locked for updation.
Possible Cause
The carbon.properties file is not identical in all the nodes of the cluster.
Procedure
Follow the steps to ensure the carbon.properties file is consistent across all the nodes:
Copy the carbon.properties file from the master node to all the other nodes in the cluster. For example, you can use scp to copy this file to all the nodes.
For the changes to take effect, restart the Spark cluster.

18)Failed to connect to hiveuser with thrift

Answer)Symptom
We get the following exception :
Cannot connect to hiveuser.
Possible Cause
The external process does not have permission to access.
Procedure
Ensure that the Hiveuser in mysql must allow its access to the external processes.

18)Failed to read the metastore db during table creation

Answer)Symptom
We get the following exception on trying to connect :
Cannot read the metastore db
Possible Cause
The metastore db is dysfunctional.
Procedure
Remove the metastore db from the carbon.metastore in the Spark Directory.

19)Failed to load data on the cluster

Answer)Symptom
Data loading fails with the following exception :
Data Load failure exception
Possible Cause
The following issue can cause the failure :
The core-site.xml, hive-site.xml, yarn-site and carbon.properties are not consistent across all nodes of the cluster.
Path to hdfs ddl is not configured correctly in the carbon.properties.
Procedure
Follow the steps to ensure the following configuration files are consistent across all the nodes:
Copy the core-site.xml, hive-site.xml, yarn-site,carbon.properties files from the master node to all the other nodes in the cluster. For example, you can use scp to copy this file to all the nodes.

www.smartdatacamp.com

Note : Set the path to hdfs ddl in carbon.properties in the master node.
For the changes to take effect, restart the Spark cluster.

19)Failed to insert data on the cluster

Answer)Symptom
Insertion fails with the following exception :
Data Load failure exception
Possible Cause
The following issue can cause the failure :
The core-site.xml, hive-site.xml, yarn-site and carbon.properties are not consistent across all nodes of the cluster.
Path to hdfs ddl is not configured correctly in the carbon.properties.
Procedure
Follow the steps to ensure the following configuration files are consistent across all the nodes:

Copy the core-site.xml, hive-site.xml, yarn-site,carbon.properties files from the master node to all the other nodes in the cluster. For example, you can use scp to copy this file to all the nodes.
Note : Set the path to hdfs ddl in carbon.properties in the master node.
For the changes to take effect, restart the Spark cluster.

20)Failed to execute Concurrent Operations on table by multiple workers

Answer)Symptom
Execution fails with the following exception :
Table is locked for updation.
Possible Cause
Concurrency not supported.
Procedure
Worker must wait for the query execution to complete and the table to release the lock for another query execution to succeed.

21)Failed to create a table with a single numeric column

Answer)Symptom
Execution fails with the following exception :
Table creation fails.
Possible Cause
Behaviour not supported.
Procedure
A single column that can be considered as dimension is mandatory for table creation.

22)Failed to create datamap and drop datamap is also not working

Answer)Symptom
Execution fails with the following exception :
HDFS Quota Exceeded
Possible Cause
HDFS Quota is set, and it is not letting carbondata write or modify any files.
Procedure
Drop that particular datamap using Drop Table command using table name as parentTableName_datamapName so as to clear the stale folders.

www.smartdatacamp.com

Apache Daffodil

Apache Daffodil is a library, requiring Java 8, used to convert between fixed format data and XML/JSON based on a DFDL schema. Some examples show the result of Daffodil parsing various inputs into XML.

1) When should I use an XSD facet like maxLength, and when should I use the DFDL length property?

Answer)Here's part of an example from the DFDL tutorial of a street address:

<xs:element name="houseNumber" type="xs:string" dfdl:lengthKind="explicit" dfdl:length="6"/>
Note that the length of the house number is constrained with DFDL. XSD can also be used to constrain lengths.
When should you used XSD to do this, and when should you use DFDL? Should you ever use both?
You must use the dfdl:length property, because it can't parse the data without it. You may use the XSD facets to check further, and it often makes sense to use both.
Consider
<xs:element name="article" type="xs:string" dfdl:length="{ ../header/articleLength }" dfdl:lengthKind='explicit'/>
Now the length is coming from a field someplace at runtime. Validating that it is within some additional constraints on maxLength might be very valuable. To do that you nave to write the more verbose:
<xs:element name="article" dfdl:length="{ ../header/articleLength }" dfdl:lengthKind='explicit'>
<xs:simpleType>
<xs:restriction base="xs:string">
<xs:maxLength value="140"/>
</xs:restriction>
</xs:simpleType>
</xs:element>
Not too bad actually. And if you can reuse some simple type definitions it's not bad at all.

One further point. Suppose you want to parse the string using the header-supplied length, but it's flat out a parse error if the length turns out to be greater than 140. You can ask the DFDL processor to check the facet maxLength at parse time using an assertion like this:
<xs:element name="article" dfdl:length="{ ../header/articleLength }" dfdl:lengthKind='explicit'>
<xs:simpleType>
<xs:annotation><xs:appinfo source="http://www.ogf.org/dfdl/dfdl-1.0">
<dfdl:assert>{ dfdl:checkConstraints() }</dfdl:assert>
</xs:appinfo></xs:annotation>
<xs:restriction base="xs:string">
<xs:maxLength value="140"/>
</xs:restriction>

</xs:simpleType>
</xs:element>
The dfdl:assert statement annotation calls a built-in DFDL function called dfdl:checkConstraints, which tells DFDL to test the facet constraints and issue a parse error if they are not satisfied. This is particularly useful for enumeration constraints where an element value is an identifier of some sort.

2)Should I use dfdl:assert to validate while parsing?

Answer)In general, no. The dfdl:assert statement annotation should be used to guide the parser. It should test things that must be true in order to successfully parse the data and create an Infoset from it.
But, it should not be used to ensure validation of the values of the data elements.
By way of illustrating what not to do, it is tempting to put facet constraints on simple type definitions in your schema, and then use a dfdl:assert like this:
<dfdl:assert>{ checkConstraints(.) }</dfdl:assert>
so that the parser will validate as it parses, and will fail to parse values that do not satisfy the facet constraints.
Don't do this. Your schema will not be as useful because it will not be able to be used for some applications, for example, applications that want to accept well-formed, but invalid data and analyze, act, or report on the invalid aspects.
In some sense, embedding checks like this into a DFDL schema is second-guessing the application's needs, and assuming the application does not even want to successfully parse and create an infoset from data that does not obey the facet constraints.

3)How do I prevent my DFDL expressions and regular expressions from being modified by my XML editor?

Answer)Use CDATA with expressions and regular expressions, and generally to stop XML editors from messing with your DFDL schema layouts.
Most XML editors will wrap long lines. So your
<a>foobar
just might get turned into
<a>foobar
Now most of the time that is fine. But sometimes the whitespace really matters. One such place is when you type a regular expression. In DFDL this can come up in this way:
<dfdl:assert testKind="pattern"> *</dfdl:assert>
Now the contents of that element is " *", i.e., a single space, and the "*" character. That means zero or more spaces in regex language. If you don't want your XML tooling to mess with the whitespace do this instead:
<dfdl:assert testKind="pattern"><![CDATA[*]]></dfdl:assert>
CDATA informs XML processors that you very much care about this. Any decent XML tooling/editor will see this and decide it cannot line-wrap this or in any way mess with the whitespace. Also useful if you want to write a complex DFDL expression in the expression language, and you want indentation and lines to be respected. Here's an example:

www.smartdatacamp.com

```
<dfdl:discriminator><![CDATA[{
if (daf:trace((daf:trace(../../ex:presenceBit,"presenceBit") = 0),"pbIsZero")) then false()
else if
(daf:trace(daf:trace(dfdl:occursIndex(),"occursIndex") = 1,"indexIsOne")) then true()
else if
(daf:trace(daf:trace(xs:int(daf:trace(../../ex:A1[daf:trace(dfdl:occursIndex()-1,"indexMinusOne")],
"occursIndexMinusOneNode")/ex:repeatBit),
"priorRepeatBit") = 0,
"priorRepeatBitIsZero"))
then false()
else true()
}]]></dfdl:discriminator>
```

If you get done writing something very deeply nested like this (and XPath style languages require this all the time), then you do NOT want anything messing with the whitespace.

About the xml:space='preserve' attribute: According to this page, xml:space is only about whitespace-only nodes, not nodes that are part whitespace. Within element-only content, the text nodes found between the elements are whitespace-only nodes. Unless you use xml:space='preserve', those are eliminated. None of the above discussion is about whitespace-only nodes. It's about value nodes containing text strings with surrounding whitespace.

4)Why doesn't DFDL allow me to express my format using offsets into a file, instead of lengths?

Answer)With some study, the DFDL workgroup concluded that these formats nearly always require the full complexity of a transformation system AND a data format description system. DFDL is only about the latter problem.

In other words, it was left out for complexity reasons, not because we didn't think there were examples.

It is a much more complex issue than people think. As we got into it we kept falling down the slippery slope of needing rich transformations to express such things.

We certainly have seen formats where there are a bunch of fields, in the ordinary manner, but instead of expressing their lengths, the forrmat specifies only their starting positions relative to start of record. There are also formats where there are tables of offsets into a subsequent data array.

DFDL requires one to recast such a specification as lengths.

It is not a "either or" scenario where lengths and offsets are equivalent so you can pick one.

Use of lengths is simply a superior and more precise way of expressing the format because use of offsets can obscure aliasing, which is the term for when there are two elements (or more) that describe the same part of the data representation. With lengths, it's clear what every bit means, and that every bit is in fact described or explicitly skipped. You can't just use an offset to skip past a bunch of data leaving it not described at all. You can't have aliasing of the same data.

Aliasing is a difficult issue when parsing. When unparsing it is a nightmare, as it introduces non-determinacy in what the data written actually comes out like. It depends on who writes it last with what alias.

Structures like

<offset to start><length of thing>
<offset to start2><length of thing2>
...
<offset to startN><length of thingN>
thing
thing2
...
thingN

So long as the things and the corresponding descriptor pairs are in order, these can be described. The lengths need not even be there as they are redundant. If present they can be checked for validity. Overlap can be checked for and deemed invalid.

But, in DFDL the above must be represented as two vectors. One of the offsets table, the other of the things. If you want an array of things and then want DFDL to convert that into the offsets and things separately, well DFDL doesn't do transformations of that sort. Do that first in XSLT or other transformation system when unparsing. When parsing, you first parse with DFDL, then transform the data into the logical single vector using XSLT (or other).

XProc is a language for expressing chains of XML-oriented transformations like this. Calabash is an open-source XProc implementation, and the daffodil-calabash-extension provides Daffodil stages that have been created to enable creation of XProc pipelines that glue together transformations like XSLT with DFDL parse/unparse steps. This can be used to create a unit that runs both DFDL and an XSLT together for parse or for unparse (they would be different XSLTs). If the things are potentially out of order, especially if the lengths are not stored, but just implied by "from this offset to the start of the next one, whichever one that is", that is simply too complex a transformation for DFDL.

If you think about what is required mentally to decode this efficiently, you must grab all the entries, sort them by offset, and then compute lengths, etc. Shy of building a real programming language (e.g., XQuery) into DFDL there has to be a limit to what level of complexity we allow DFDL to express directly. And unparsing is entirely non-deterministic... you have to stage an array/blob filled with fill bytes, write pieces to it one by one, potentially overwriting sections. It's really quite hard. Even if you supported this in DFDL somehow, would it in fact write these things out in the order an application does? So will you even be able to re-create data?

There is a sense in which formats expressed as these sorts of "potentially overlapping regions" are simply not adequately specified unless they specify the exact order things are to be written so that the contents of overlap regions is deterministic.

There could be formats where there are offset tables like this, where in principle things could be out of order, or overlapping/aliased, but they simply never are, and allowing them to be is effectively a bad idea as it allows people to do very obscure things - information hiding, polyglot files, etc. PDF is heavily criticized for this. It may be an unstated principle that such formats do not do this sort of out-of-order or aliasing stuff.

All that said, practically speaking, people have data with offset tables, and out-of-order might be a possibility that needs to be allowed at least on parsing. So what to do in DFDL?

In this case, DFDL can describe the table of offsets, and a big blob of data. Beyond that something else (e.g., XSLT, or a program) must take over for expressing the sort and extraction of chunks out of the larger blob.

If you think about this, if you want deterministic unparsing behavior, that is what has to be presented to the DFDL unparser anyway, since presenting the resolved content blob means the

application has dealt with the order to which the various chunks (which may overlap) have been written.

5)How can I get strings in the data to become element names?

Answer)If the data contains tags/strings, and you want those strings to become element names in XML, then you must do pass 1 to extract the tag information, use them as element names when you create a DFDL schema dynamically, and then parse the data again with this new specialized DFDL schema.
Or you can parse the data with a generic schema where your tag names will be in element values someplace, and do a transformation outside of DFDL to convert them to element names.
Consider the common “comma separated values” or CSV formats. If you have
Name, Address, Phone
Mike, 8840 Standford Blvd\, Columbia MD, 888-888-8888
and you want
<columnNames>
<name>Name</name>
<name>Address</name>
<name>Phone</name>
</columnNames>
<row>
<col>Mike</col>
<col>8840 Standford Blvd, Columbia MD</col>
<col>888-888-8888</col>
</row>
That’s what you would get from a generic CSV DFDL schema. If you want this:
<row>
<Name>Mike</Name>
<Address>8840 Stanford Blvd, Columbia MD</Address>
<Phone>888-888-8888</Phone>
</row>
That’s a specific-to-exactly-these-column-names CSV DFDL schema that is required. If you have lots of files with this exact structure you would create this DFDL schema once.
If you have no idea what CSV is coming at you, but want this sort of XML elements anyway, then you have to generate a DFDL schema on the fly from the data (parse just the headers with a generic DFDL schema first - then use that to create the DFDL schema.
Or you parse using the generic schema, then use XSLT or something to convert the result of the generic parse.
Keep in mind that this problem has little to do with DFDL. Given an XML document like the generic one above, but you didn’t want that XML, you wanted the specific style XML. Well you have the same problem. You need to grab the column names first, then transform the data using them as the element names.

Apache Drill

Drill is an Apache open-source SQL query engine for Big Data exploration. Drill is designed from the ground up to support high-performance analysis on the semi-structured and rapidly evolving data coming from modern Big Data applications, while still providing the familiarity and ecosystem of ANSI SQL, the industry-standard query language. Drill provides plug-and-play integration with existing Apache Hive and Apache HBase deployments.

1) Why Drill?

Answer)The 40-year monopoly of the RDBMS is over. With the exponential growth of data in recent years, and the shift towards rapid application development, new data is increasingly being stored in non-relational datastores including Hadoop, NoSQL and cloud storage. Apache Drill enables analysts, business users, data scientists and developers to explore and analyze this data without sacrificing the flexibility and agility offered by these datastores. Drill processes the data in-situ without requiring users to define schemas or transform data.

2)What are some of Drill's key features?

Answer)Drill is an innovative distributed SQL engine designed to enable data exploration and analytics on non-relational datastores. Users can query the data using standard SQL and BI tools without having to create and manage schemas. Some of the key features are:
Schema-free JSON document model similar to MongoDB and Elasticsearch
Industry-standard APIs: ANSI SQL, ODBC/JDBC, RESTful APIs
Extremely user and developer friendly
Pluggable architecture enables connectivity to multiple datastores

3)How does Drill achieve performance?

Answer)Drill is built from the ground up to achieve high throughput and low latency. The following capabilities help accomplish that:
Distributed query optimization and execution: Drill is designed to scale from a single node (your laptop) to large clusters with thousands of servers.
Columnar execution: Drill is the world's only columnar execution engine that supports complex data and schema-free data. It uses a shredded, in-memory, columnar data representation.
Runtime compilation and code generation: Drill is the world's only query engine that compiles and re-compiles queries at runtime. This allows Drill to achieve high performance without knowing the structure of the data in advance. Drill leverages multiple compilers as well as ASM-based bytecode rewriting to optimize the code.

www.smartdatacamp.com

Vectorization: Drill takes advantage of the latest SIMD instructions available in modern processors.
Optimistic/pipelined execution: Drill is able to stream data in memory between operators. Drill minimizes the use of disks unless needed to complete the query.

4)What datastores does Drill support?

Answer)Drill is primarily focused on non-relational datastores, including Hadoop, NoSQL and cloud storage. The following datastores are currently supported:
Hadoop: All Hadoop distributions (HDFS API 2.3+), including Apache Hadoop, MapR, CDH and Amazon EMR
NoSQL: MongoDB, HBase
Cloud storage: Amazon S3, Google Cloud Storage, Azure Blog Storage, Swift
A new datastore can be added by developing a storage plugin. Drill's unique schema-free JSON data model enables it to query non-relational datastores in-situ (many of these systems store complex or schema-free data).

5)What clients are supported?

Answer)BI tools via the ODBC and JDBC drivers (eg, Tableau, Excel, MicroStrategy, Spotfire, QlikView, Business Objects)
Custom applications via the REST API
Java and C applications via the dedicated Java and C libraries

6)Is Drill a SQL-on-Hadoop engine?

Answer)Drill supports a variety of non-relational datastores in addition to Hadoop. Drill takes a different approach compared to traditional SQL-on-Hadoop technologies like Hive and Impala. For example, users can directly query self-describing data (eg, JSON, Parquet) without having to create and manage schemas.
The following table provides a more detailed comparison between Drill and traditional SQL-on-Hadoop technologies:
| Drill | SQL-on-Hadoop (Hive, Impala, etc.)
Use case | Self-service, in-situ, SQL-based analytics | Data warehouse offload
Data sources | Hadoop, NoSQL, cloud storage (including multiple instances) | A single Hadoop cluster
Data model | Schema-free JSON (like MongoDB) | Relational
User experience | Point-and-query | Ingest data → define schemas → query
Deployment model| Standalone service or co-located with Hadoop or NoSQL | Co-located with Hadoop
Data management | Self-service | IT-driven
SQL | ANSI SQL | SQL-like
1.0 availability| Q2 2015 | Q2 2013 or earlier

7)Is Spark SQL similar to Drill?

Answer)No. Spark SQL is primarily designed to enable developers to incorporate SQL statements in Spark programs. Drill does not depend on Spark, and is targeted at business users, analysts, data scientists and developers.

8)Does Drill replace Hive?

Answer)Hive is a batch processing framework most suitable for long-running jobs. For data exploration and BI, Drill provides a much better experience than Hive.

In addition, Drill is not limited to Hadoop. For example, it can query NoSQL databases (eg, MongoDB, HBase) and cloud storage (eg, Amazon S3, Google Cloud Storage, Azure Blob Storage, Swift).

9)How does Drill support queries on self-describing data?

Answer)Drill's flexible JSON data model and on-the-fly schema discovery enable it to query self-describing data.

JSON data model: Traditional query engines have a relational data model, which is limited to flat records with a fixed structure. Drill is built from the ground up to support modern complex/semi-structured data commonly seen in non-relational datastores such as Hadoop, NoSQL and cloud storage. Drill's internal in-memory data representation is hierarchical and columnar, allowing it to perform efficient SQL processing on complex data without flattening into rows.

On-the-fly schema discovery (or late binding): Traditional query engines (eg, relational databases, Hive, Impala, Spark SQL) need to know the structure of the data before query execution. Drill, on the other hand, features a fundamentally different architecture, which enables execution to begin without knowing the structure of the data. The query is automatically compiled and re-compiled during the execution phase, based on the actual data flowing through the system. As a result, Drill can handle data with evolving schema or even no schema at all (eg, JSON files, MongoDB collections, HBase tables).

10)But I already have schemas defined in Hive Metastore? Can I use that with Drill?

Answer)Absolutely. Drill has a storage plugin for Hive tables, so you can simply point Drill to the Hive Metastore and start performing low-latency queries on Hive tables. In fact, a single Drill cluster can query data from multiple Hive Metastores, and even perform joins across these datasets.

www.smartdatacamp.com

11)Is Drill "anti-schema" or "anti-DBA"?

Answer)Not at all. Drill actually takes advantage of schemas when available. For example, Drill leverages the schema information in Hive when querying Hive tables. However, when querying schema-free datastores like MongoDB, or raw files on S3 or Hadoop, schemas are not available, and Drill is still able to query that data.

Centralized schemas work well if the data structure is static, and the value of data is well understood and ready to be operationalized for regular reporting purposes. However, during data exploration, discovery and interactive analysis, requiring rigid modeling poses significant challenges. For example:

Complex data (eg, JSON) is hard to map to relational tables

Centralized schemas are hard to keep in sync when the data structure is changing rapidly

Non-repetitive/ad-hoc queries and data exploration needs may not justify modeling costs

Drill is all about flexibility. The flexible schema management capabilities in Drill allow users to explore raw data and then create models/structure with CREATE TABLE or CREATE VIEW statements, or with Hive Metastore.

11)What does a Drill query look like?

Answer)Drill uses a decentralized metadata model and relies on its storage plugins to provide metadata. There is a storage plugin associated with each data source that is supported by Drill.

The name of the table in a query tells Drill where to get the data:
SELECT * FROM dfs1.root.`/my/log/files/`;
SELECT * FROM dfs2.root.`/home/john/log.json`;
SELECT * FROM mongodb1.website.users;
SELECT * FROM hive1.logs.frontend;
SELECT * FROM hbase1.events.clicks;

12)What SQL functionality does Drill support?

Answer)Drill supports standard SQL (aka ANSI SQL). In addition, it features several extensions that help with complex data, such as the KVGEN and FLATTEN functions.

13)Do I need to load data into Drill to start querying it?

Answer)No. Drill can query data 'in-situ'.

Apache Edgent

Apache Edgent is a programming model and micro-kernel style runtime that can be embedded in gateways and small footprint edge devices enabling local, real-time, analytics on the continuous streams of data coming from equipment, vehicles, systems, appliances, devices and sensors of all kinds (for example, Raspberry Pis or smart phones). Working in conjunction with centralized analytic systems, Apache Edgent provides efficient and timely analytics across the whole IoT ecosystem: from the center to the edge.

1) What is Apache Edgent?

Answer)Edgent provides APIs and a lightweight runtime enabling you to easily create event-driven flow-graph style applications to analyze streaming data at the edge.

2)What do you mean by the edge?

Answer)The edge includes devices, gateways, equipment, vehicles, systems, appliances and sensors of all kinds as part of the Internet of Things.
It's easy for for Edgent applications to connect to other entities such as an enterprise IoT hub. While Edgent's design center is executing on constrained edge devices, Edgent applications can run on any system meeting minimal requirements such as a Java runtime.

3)How are applications developed?

Answer)Applications are developed using a functional flow API to define operations on data streams that are executed as a flow graph in a lightweight embeddable runtime. Edgent provides capabilities like windowing, aggregation and connectors with an extensible model for the community to expand its capabilities. Check out The Power of Edgent!

You can develop Edgent applications using an IDE of your choice.

Generally, mechanisms for deploying an Edgent application to a device are beyond the scope of Edgent; they are often device specific or may be defined by an enterprise IoT system. To deploy an Edgent application to a device like a Raspberry Pi, you could just FTP the application to the device and modify the device to start the application upon startup or on command. See Edgent application Development.

4)What environments does Apache Edgent support?

www.smartdatacamp.com

Answer)Currently, Edgent provides APIs and runtime for Java and Android. Support for additional languages, such as Python, is likely as more developers get involved. Please consider joining the Edgent open source development community to accelerate the contributions of additional APIs.

5)What type of analytics can be done with Apache Edgent?

Answer)The core Edgent APIs make it easy to incorporate any analytics you want into the stream processing graph. It's trivial to create windows and trigger aggregation functions you supply. It's trivial to specify whatever filtering and transformation functions you want to supply. The functions you supply can use existing libraries.

Edgent comes with some initial analytics for aggregation and filtering that you may find useful. It uses Apache Common Math to provide simple analytics aimed at device sensors. In the future, Edgent will include more analytics, either exposing more functionality from Apache Common Math, other libraries or hand-coded analytics.

6)What connectors does Apache Edgent support?

Answer)Edgent provides easy to use connectors for MQTT, HTTP, JDBC, File, Apache Kafka and IBM Watson IoT Platform. Edgent is extensible; you can create connectors. You can easily supply any code you want for ingesting data from and sinking data to external systems.

7)Does Edgent have a Sensor Library?

Answer)No, Edgent does not come with a library for accessing a device's sensors. The simplicity with which an Edgent application can poll or otherwise use existing APIs for reading a sensor value make such a library unnecessary.

8)What centralized streaming analytic systems does Apache Edgent support?

Answer)Edgent applications can publish and subscribe to message systems like MQTT or Kafka, or IoT Hubs like IBM Watson IoT Platform. Centralized streaming analytic systems can do likewise to then consume Edgent application events and data, as well as control an Edgent application. The centralized streaming analytic system could be Apache Spark, Apache Storm, Flink and Samza, IBM Streams (on-premises or IBM Streaming Analytics on Bluemix), or any custom application of your choice.

9)Is there a distributed version of Edgent?

Answer)The short answer is that a single Edgent application's topologies all run in the same local JVM.

But sometimes this question is really asking "Can separate Edgent topologies communicate with each other?" and the answer to that is YES!

Today, multiple topologies in a single Edgent application/JVM can communicate using the Edgent PublishSubscribe connector, or any other shared resource you choose to use (e.g., a java.util.concurrent.BlockingQueue).

Edgent topologies in separate JVM's, or the same JVM, can communicate with each other by using existing connectors to a local or remote MQTT server for example.

10)Why do I need Apache Edgent on the edge, rather than my streaming analytic system?

Answer)Edgent is designed for the edge. It has a small footprint, suitable for running on constrained devices. Edgent applications can analyze data on the edge and to only send to the centralized system if there is a need, reducing communication costs.

11)Why do I need Apache Edgent, rather than coding the complete application myself?

Answer)Edgent is designed to accelerate your development of edge analytic applications - to make you more productive! Edgent provides a simple yet powerful consistent data model (streams and windows) and provides useful functionality, such as aggregations, joins, and connectors. Using Edgent lets you to take advantage of this functionality, allowing you to focus on your application needs.

11)Why is Apache Edgent open source?

Answer)With the growth of the Internet of Things there is a need to execute analytics at the edge. Edgent was developed to address requirements for analytics at the edge for IoT use cases that were not addressed by central analytic solutions. These capabilities will be useful to many organizations and that the diverse nature of edge devices and use cases is best addressed by an open community. Our goal is to develop a vibrant community of developers and users to expand the capabilities and real-world use of Edgent by companies and individuals to enable edge analytics and further innovation for the IoT space.

www.smartdatacamp.com

Apache Flink

Apache Flink is a framework and distributed processing engine for stateful computations over unbounded and bounded data streams. Flink has been designed to run in all common cluster environments, perform computations at in-memory speed and at any scale.

1)Is Apache Flink only for (near) real-time processing use cases?

Answer)Flink is a very general system for data processing and data-driven applications with data streams as the core building block. These data streams can be streams of real-time data, or stored streams of historic data. For example, in Flink's view a file is a stored stream of bytes. Because of that, Flink supports both real-time data processing and applications, as well as batch processing applications.

Streams can be unbounded (have no end, events continuously keep coming) or be bounded (streams have a beginning and an end). For example, a Twitter feed or a stream of events from a message queue are generally unbounded streams, whereas a stream of bytes from a file is a bounded stream.

2)If everything is a stream, why are there a DataStream and a DataSet API in Flink?

Answer)Bounded streams are often more efficient to process than unbounded streams. Processing unbounded streams of events in (near) real-time requires the system to be able to immediately act on events and to produce intermediate results (often with low latency). Processing bounded streams usually does not require producing low latency results, because the data is a while old anyway (in relative terms). That allows Flink to process the data in a simple and more efficient way.

The DataStream API captures the continuous processing of unbounded and bounded streams, with a model that supports low latency results and flexible reaction to events and time (including event time).

The DataSet API has techniques that often speed up the processing of bounded data streams. In the future, the community plans to combine these optimizations with the techniques in the DataStream API.

3)How does Flink relate to the Hadoop Stack?

Answer)Flink is independent of Apache Hadoop and runs without any Hadoop dependencies.

However, Flink integrates very well with many Hadoop components, for example, HDFS, YARN, or HBase. When running together with these components, Flink can use HDFS to read data, or write

results and checkpoints/snapshots. Flink can be easily deployed via YARN and integrates with the YARN and HDFS Kerberos security modules.

4)What other stacks does Flink run in?

Answer)Users run Flink on Kubernetes, Mesos, Docker, or even as standalone services.

5)What are the prerequisites to use Flink?

Answer)You need Java 8 to run Flink jobs/applications.

The Scala API (optional) depends on Scala 2.11.

Highly-available setups with no single point of failure require Apache ZooKeeper.

For highly-available stream processing setups that can recover from failures, Flink requires some form of distributed storage for checkpoints (HDFS / S3 / NFS / SAN / GFS / Kosmos / Ceph / ...).

6)What scale does Flink support?

Answer)Users are running Flink jobs both in very small setups (fewer than 5 nodes) and on 1000s of nodes and with TBs of state.

7)Is Flink limited to in-memory data sets?

Answer)For the DataStream API, Flink supports larger-than-memory state be configuring the RocksDB state backend.

For the DataSet API, all operations (except delta-iterations) can scale beyond main memory.

www.smartdatacamp.com

Apache Hama

Apache Hama is a framework for Big Data analytics which uses the Bulk Synchronous Parallel (BSP) computing model, which was established in 2012 as a Top-Level Project of The Apache Software Foundation.

It provides not only pure BSP programming model but also vertex and neuron centric programming models, inspired by Google's Pregel and DistBelief.

1) I get ": hostname nor servname provided, or not known" error on Cygwin/Windows.

Answer)You can fix this by changing hama.zookeeper.quorum variable 'localhost' to '127.0.0.1'.

2) I get ": Incorrect header or version mismatch from 127.0.0.1:52772 got version 3 expected version 4." while starting

Answer)Please use a release of Hadoop that is compatible with the Hama Release.

3)I get ": FATAL org.apache.hama.BSPMasterRunner: java.net.UnknownHostException: Invalid hostname for server: local" while starting.

Answer) This is the case if you're in the local-mode and tried to launch Hama via the start script. In this mode, nothing has to be launched. A multithreaded running utility will start when submitting your job.

4)When I submit a job, I see that it fails immediately without running a task.

Answer)Please look in your BSPMaster.log in the log directory under $HAMA_HOME/logs/hama-$USER-bspmaster-$HOSTNAME.log. If you see a line equal to

2012-07-28 17:45:34,708 ERROR org.apache.hama.bsp.SimpleTaskScheduler: Scheduling of job test.jar could not be done successfully. Killing it!

the scheduler could not schedule your job, because you don't have enough resources (task slots) in your cluster available. So watch closely while submitting the job, if it says

2012-07-28 17:45:34 INFO bsp.FileInputFormat: Total # of splits: 4

and your cluster shows (for example in the web UI) only 3 slots that are free, our scheduler could not successfully schedule all the tasks. If you are familiar with Hadoop, you will be confused with this behaviour. Mainly because BSP needs the tasks to run in parallel, whereas in MapReduce the

map tasks are not depending on each other (so they can be processed after each other). We are sorry for the not existing error message and will fix this in near future.

5)Is there any restriction on max message sent before sync()?

Answer)In Mem-based queue case, messages is kept in memory, therefore it depends on memory available. In Spilling queue case, there's no limits.

6) java.lang.IllegalArgumentException: Messages must never be behind the vertex in ID!

Answer)This exception will be throwed out when received message belongs to non-existent vertex (dangling links). To ignore them, set "hama.check.missing.vertex" to false.

www.smartdatacamp.com

SQL

SQL or Structured Query Language used in programming and designed for managing data held in a relational database management system (RDBMS)

SQL offers two main advantages over older read–write APIs such as ISAM or VSAM. Firstly, it introduced the concept of accessing many records with one single command. Secondly, it eliminates the need to specify how to reach a record, e.g. with or without an index.

1) What is the SQL server query execution sequence?

Answer) FROM -> goes to Secondary files via primary file.
WHERE -> applies filter condition (non-aggregate column)
SELECT -> dumps data in tempDB system database
GROUP BY -> groups data according to grouping predicate
HAVING -> applies filter condition (aggregate function)
ORDER BY -> sorts data ascending/descending

2) What is Normalization?

Answer)Step by step process to reduce the degree of data redundancy.
Breaking down one big flat table into multiple table based on normalization rules.
Optimizing the memory but not in term of performance.
Normalization will get rid of insert, update and delete anomalies.
Normalization will improve the performance of the delta operation (aka. DML operation); UPDATE, INSERT, DELETE
Normalization will reduce the performance of the read operation; SELECT

3)What are the three degrees of normalization and how is normalization done in each degree?.

Answer) 1NF:
A table is in 1NF when: All the attributes are single-valued.
With no repeating columns (in other words, there cannot be two different columns with the same information).
With no repeating rows (in other words, the table must have a primary key).
All the composite attributes are broken down into its minimal component.
There should be SOME (full, partial, or transitive) kind of functional dependencies between non-key and key attributes.
99% of times, it's usually 1NF.
2NF:
A table is in 2NF when: It is in 1NF.

There should not be any partial dependencies so they must be removed if they exist.
3NF:
A table is in 3NF when: It is in 2NF.
There should not be any transitive dependencies so they must be removed if they exist.
BCNF:
A stronger form of 3NF so it is also known as 3.5NF
We do not need to know much about it. Just know that here you compare between a prime attribute and a prime attribute and a non-key attribute and a non-key attribute.

4)What are the different database objects ?

Answer)There are total seven database objects (6 permanent database object + 1 temporary database object)
Permanent DB objects
Table
Views
Stored procedures
User-defined Functions
Triggers
Indexes
Temporary DB object
Cursors

5)What is collation?

Answer)Bigdata Hadoop: SQL Interview Question with Answers
Collation is defined as set of rules that determine how character data can be sorted and compared.
This can be used to compare A and, other language characters and also depends on the width of the characters.
ASCII value can be used to compare these character data.

6)What is a constraint and what are the seven constraints?

Answer)Constraint: something that limits the flow in a database.
1. Primary key
2. Foreign key
3. Check
Ex: check if the salary of employees is over 40,000
4. Default
Ex: If the salary of an employee is missing, place it with the default value.
5. Nullability
NULL or NOT NULL
6. Unique Key

www.smartdatacamp.com

7. Surrogate Key
mainly used in data warehouse

7)What is a Surrogate Key ?

Answer)Surrogate means Substitute.

Surrogate key is always implemented with a help of an identity column.

Identity column is a column in which the value are automatically generated by a SQL Server based on the seed value and incremental value.

Identity columns are ALWAYS INT, which means surrogate keys must be INT. Identity columns cannot have any NULL and cannot have repeated values. Surrogate key is a logical key.

8)What is a derived column , hows does it work , how it affects the performance of a database and how can it be improved?

Answer)The Derived Column a new column that is generated on the fly by applying expressions to transformation input columns.

Ex: FirstName + ' ' + LastName AS 'Full name'

Derived column affect the performances of the data base due to the creation of a temporary new column.

Execution plan can save the new column to have better performance next time.

9)What is a Transaction?

Answer)It is a set of TSQL statement that must be executed together as a single logical unit.

Has ACID properties:

Atomicity: Transactions on the DB should be all or nothing. So transactions make sure that any operations in the transaction happen or none of them do.

Consistency: Values inside the DB should be consistent with the constraints and integrity of the DB before and after a transaction has completed or failed.

Isolation: Ensures that each transaction is separated from any other transaction occurring on the system.

Durability: After successfully being committed to the RDMBS system the transaction will not be lost in the event of a system failure or error.

Actions performed on explicit transaction:

BEGIN TRANSACTION: marks the starting point of an explicit transaction for a connection.

COMMIT TRANSACTION (transaction ends): used to end an transaction successfully if no errors were encountered. All DML changes made in the transaction become permanent.

ROLLBACK TRANSACTION (transaction ends): used to erase a transaction which errors are encountered. All DML changes made in the transaction are undone.

SAVE TRANSACTION (transaction is still active): sets a savepoint in a transaction. If we roll back, we can only rollback to the most recent savepoint. Only one save point is possible per transaction. However, if you nest Transactions within a Master Trans, you may put Save points in each nested Tran. That is how you create more than one Save point in a Master Transaction.

10)What are the differences between OLTP and OLAP?

Answer)OLTP stands for Online Transactional Processing
OLAP stands for Online Analytical Processing
OLTP:
Normalization Level: highly normalized
Data Usage : Current Data (Database)
Processing : fast for delta operations (DML)
Operation : Delta operation (update, insert, delete) aka DML Terms Used : table, columns and relationships
OLAP:
Normalization Level: highly denormalized
Data Usage : historical Data (Data warehouse)
Processing : fast for read operations
Operation : read operation (select)
Terms Used : dimension table, fact table

11)How do you copy just the structure of a table?

Answer)SELECT * INTO NewDB.TBL_Structure
FROM OldDB.TBL_Structure
WHERE 1=0 -- Put any condition that does not make any sense.

12)What are the different types of Joins?

Answer) INNER JOIN: Gets all the matching records from both the left and right tables based on joining columns.
LEFT OUTER JOIN: Gets all non-matching records from left table & AND one copy of matching records from both the tables based on the joining columns.
RIGHT OUTER JOIN: Gets all non-matching records from right table & AND one copy of matching records from both the tables based on the joining columns.
FULL OUTER JOIN: Gets all non-matching records from left table & all non-matching records from right table & one copy of matching records from both the tables.

www.smartdatacamp.com

CROSS JOIN: returns the Cartesian product.

13)What are the different types of Restricted Joins?

Answer)SELF JOIN: joining a table to itself
RESTRICTED LEFT OUTER JOIN: gets all non-matching records from left side
RESTRICTED RIGHT OUTER JOIN - gets all non-matching records from right side
RESTRICTED FULL OUTER JOIN - gets all non-matching records from left table & gets all nonmatching records from right table.

14)What is a sub-query?

Answer)It is a query within a query
Syntax:
SELECT >column_name< FROM >table_name<
WHERE >column_name< IN/NOT IN
(
>another SELECT statement<
)
Everything that we can do using sub queries can be done using Joins, but anything that we can do using Joins may/may not be done using Subquery.
Sub-Query consists of an inner query and outer query. Inner query is a SELECT statement the result of which is passed to the outer query. The outer query can be SELECT, UPDATE, DELETE. The result of the inner query is generally used to filter what we select from the outer query We can also have a subquery inside of another subquery and so on. This is called a nested Subquery. Maximum one can have is 32 levels of nested Sub-Queries

15)What are the SET Operators?

Answer)SQL set operators allows you to combine results from two or more SELECT statements.
Syntax:
SELECT Col1, Col2, Col3 FROM T1 >SET OPERATOR<
SELECT Col1, Col2, Col3 FROM T2
Rule 1: The number of columns in first SELECT statement must be same as the number of columns in the second SELECT statement.
Rule 2: The metadata of all the columns in first SELECT statement MUST be exactly same as the metadata of all the columns in second SELECT statement accordingly.
Rule 3: ORDER BY clause do not work with first SELECT statement. ○ UNION, UNION ALL, INTERSECT, EXCEPT

16)What is a derived table?

Answer)SELECT statement that is given an alias name and can now be treated as a virtual table and operations like joins, aggregations, etc. can be performed on it like on an actual table. Scope is query bound, that is a derived table exists only in the query in which it was defined.
SELECT temp1.SalesOrderID, temp1.TotalDue FROM
(SELECT TOP 3 SalesOrderID, TotalDue FROM Sales.SalesOrderHeader ORDER BY TotalDue DESC) AS temp1 LEFT OUTER JOIN
(SELECT TOP 2 SalesOrderID, TotalDue FROM Sales.SalesOrderHeader ORDER BY TotalDue DESC) AS temp2 ON temp1.SalesOrderID = temp2.SalesOrderID WHERE temp2.SalesOrderID IS NULL

17)What is a View?

Answer)Views are database objects which are virtual tables whose structure is defined by underlying SELECT statement and is mainly used to implement security at rows and columns levels on the base table.

One can create a view on top of other views. View just needs a result set (SELECT statement).

We use views just like regular tables when it comes to query writing. (joins, subqueries, grouping)

We can perform DML operations (INSERT, DELETE, UPDATE) on a view. It actually affects the underlying tables only those columns can be affected which are visible in the view.

18)What are the types of views?

Answer)1. Regular View: It is a type of view in which you are free to make any DDL changes on the underlying table.

create a regular view

CREATE VIEW v_regular AS SELECT * FROM T1

2. Schemabinding View:

It is a type of view in which the schema of the view (column) are physically bound to the schema of the underlying table. We are not allowed to perform any DDL changes to the underlying table for the columns that are referred by the schemabinding view structure.

All objects in the SELECT query of the view must be specified in two part naming conventions

(schema_name.tablename).

You cannot use * operator in the SELECT query inside the view (individually name the columns)

All rules that apply for regular view.

CREATE VIEW v_schemabound WITH SCHEMABINDING AS SELECT ID, Name

FROM dbo.T2 -- remember to use two part naming convention

3. Indexed View:

www.smartdatacamp.com

19)What is an Indexed View?

Answer)It is technically one of the types of View, not Index.

Using Indexed Views, you can have more than one clustered index on the same table if needed.

All the indexes created on a View and underlying table are shared by Query Optimizer to select the best way to execute the query.

Both the Indexed View and Base Table are always in sync at any given point.

Indexed Views cannot have NCI-H, always NCI-CI, therefore a duplicate set of the data will be created.

20)What does WITH CHECK do?

Answer) WITH CHECK is used with a VIEW.

It is used to restrict DML operations on the view according to search predicate (WHERE clause) specified creating a view.

Users cannot perform any DML operations that do not satisfy the conditions in WHERE clause while creating a view

WITH CHECK OPTION has to have a WHERE clause.

21)What is a RANKING function and what are the four RANKING functions?

Answer)Ranking functions are used to give some ranking numbers to each row in a dataset based on some ranking functionality.

Every ranking function creates a derived column which has integer value.

Different types of RANKING function:

ROW_NUMBER(): assigns an unique number based on the ordering starting with 1. Ties will be given different ranking positions.

RANK(): assigns an unique rank based on value. When the set of ties ends, the next ranking position will consider how many tied values exist and then assign the next value a new ranking with consideration the number of those previous ties. This will make the ranking position skip placement.position numbers based on how many of the same values occurred (ranking not sequential).

DENSE_RANK(): same as rank, however it will maintain its consecutive order nature regardless of ties in values; meaning if five records have a tie in the values, the next ranking will begin with the next ranking position.

Syntax:

<Ranking Function>() OVER(condition for ordering) always have to have an OVER clause

Ex:
SELECT SalesOrderID, SalesPersonID,
TotalDue,
ROW_NUMBER() OVER(ORDER BY TotalDue), RANK() OVER(ORDER BY TotalDue),
DENSE_RANK() OVER(ORDER BY TotalDue) FROM Sales.SalesOrderHeader
NTILE(n): Distributes the rows in an ordered partition into a specified number of groups

22)What is PARTITION BY?

Answer)Creates partitions within the same result set and each partition gets its own ranking. That is, the rank starts from 1 for each partition.

Ex:
SELECT *, DENSE_RANK() OVER(PARTITION BY Country ORDER BY Sales DESC) AS DenseRank
FROM SalesInfo

23)What is Temporary Table and what are the two types of it? ○ They are tables just like regular tables but the main difference is its scope.

Answer)The scope of temp tables is temporary whereas regular tables permanently reside.Temporary table are stored in tempDB.
We can do all kinds of SQL operations with temporary tables just like regular tables like JOINs, GROUPING, ADDING CONSTRAINTS, etc.
Two types of Temporary Table
Local
LocalTempTableName -- single pound sign
Only visible in the session in which they are created. It is session-bound.
Global
GlobalTempTableName -- double pound sign
Global temporary tables are visible to all sessions after they are created, and are deleted when the session in which they were created in is disconnected.
It is last logged-on user bound. In other words, a global temporary table will disappear when the last user on the session logs off.

24)Explain Variables ?

Answer)Variable is a memory space (place holder) that contains a scalar value EXCEPT table variables, which is 2D data.

Variable in SQL Server are created using DECLARE Statement. ○ Variables are BATCH-BOUND.

Variables that start with @ are user-defined variables.

25)Explain Dynamic SQL (DSQL). ?

Answer)Dynamic SQL refers to code/script which can be used to operate on different data-sets based on some dynamic values supplied by front-end applications.

It can be used to run a template SQL query against different tables/columns/conditions.

Declare variables: which makes SQL code dynamic.

Main disadvantage of D-SQL is that we are opening SQL Tool for SQL Injection attacks.

You should build the SQL script by concatenating strings and variable.

26)What is SQL Injection Attack?

Answer)Moderator's definition: when someone is able to write a code at the front end using DSQL, he/she could use malicious code to drop, delete, or manipulate the database. There is no perfect protection from it but we can check if there is certain commands such as 'DROP' or 'DELETE' are included in the command line. SQL Injection is a technique used to attack websites by inserting SQL code in web entry fields.

27)What is SELF JOIN?

Answer)JOINing a table to itself When it comes to SELF JOIN, the foreign key of a table points to its primary key.

Ex: Employee(Eid, Name, Title, Mid) Know how to implement it!!!

28)What is Correlated Subquery?

Answer)It is a type of subquery in which the inner query depends on the outer query.

This means that that the subquery is executed repeatedly, once for each row of the outer query.

In a regular subquery, inner query generates a result set that is independent of the outer query.

Ex:

```
SELECT *
FROM HumanResources.Employee E
WHERE 5000 IN (SELECT S.Bonus
FROM Sales.SalesPerson S
WHERE S.SalesPersonID = E.EmployeeID)
```

The performance of Correlated Subquery is very slow because its inner query depends on the outer query.

So the inner subquery goes through every single row of the result of the outer subquery.

29)What is the difference between Regular Subquery and Correlated Subquery?

Answer)Based on the above explanation, an inner subquery is independent from its outer subquery in Regular Subquery.

On the other hand, an inner subquery depends on its outer subquery in Correlated Subquery.

30)What are the differences between DELETE and TRUNCATE .?

Answer)Delete:
DML statement that deletes rows from a table and can also specify rows using a WHERE clause.
Logs every row deleted in the log file.
Slower since DELETE records every row that is deleted.
DELETE continues using the earlier max value of the identity column. Can have triggers on DELETE.
Truncate:
DDL statement that wipes out the entire table and you cannot delete specific rows.
Does minimal logging, minimal as not logging everything. TRUNCATE will remove the pointers that point to their pages, which are deallocated.
Faster since TRUNCATE does not record into the log file. TRUNCATE resets the identity column.
Cannot have triggers on TRUNCATE.

31)What are the three different types of Control Flow statements?

Answer)1. WHILE
2. IF-ELSE
3. CASE

32)What is Table Variable? Explain its advantages and disadvantages.?

Answer)If we want to store tabular data in the form of rows and columns into a variable then we use a table variable. It is able to store and display 2D data (rows and columns).

We cannot perform DDL (CREATE, ALTER, DROP).
Advantages:
Table variables can be faster than permanent tables.
Table variables need less locking and logging resources.
Disadvantages:
Scope of Table variables is batch bound.
Table variables cannot have constraints.
Table variables cannot have indexes.
Table variables do not generate statistics.
Cannot ALTER once declared (Again, no DDL statements).

www.smartdatacamp.com

33)What are the differences between Temporary Table and Table Variable?

Answer)Temporary Table:
It can perform both DML and DDL Statement. Session bound Scope
Syntax CREATE TABLE #temp
Have indexes
Table Variable:
Can perform only DML, but not DDL Batch bound scope
DECLARE @var TABLE(...)
Cannot have indexes

34)What is Stored Procedure (SP)?

Answer)It is one of the permanent DB objects that is precompiled set of TSQL statements that can accept and return multiple variables.

It is used to implement the complex business process/logic. In other words, it encapsulates your entire business process.

Compiler breaks query into Tokens. And passed on to query optimizer. Where execution plan is generated the very 1st time when we execute a stored procedure after creating/altering it and same execution plan is utilized for subsequent executions.

Database engine runs the machine language query and execute the code in 0's and 1's.

When a SP is created all Tsql statements that are the part of SP are pre-compiled and execution plan is stored in DB which is referred for following executions.

Explicit DDL requires recompilation of SP's.

35)What are the four types of SP?

Answer)System Stored Procedures (SP_****): built-in stored procedures that were created by Microsoft.

User Defined Stored Procedures: stored procedures that are created by users. Common naming convention (usp_****)

CLR (Common Language Runtime): stored procedures that are implemented as public static methods on a class in a Microsoft .NET Framework assembly.

Extended Stored Procedures (XP_****): stored procedures that can be used in other platforms such as Java or C++.

36)Explain the Types of SP..? ○ SP with no parameters:

Answer)SP with a single input parameter:

SP with multiple parameters:

SP with output parameters:

Extracting data from a stored procedure based on an input parameter and outputting them using output variables.

SP with RETURN statement (the return value is always single and integer value)

37)What are the characteristics of SP?

Answer)SP can have any kind of DML and DDL statements.
SP can have error handling (TRY ...CATCH).
SP can use all types of table.
SP can output multiple integer values using OUT parameters, but can return only one scalar INT value.
SP can take any input except a table variable.
SP can set default inputs.
SP can use DSQL.
SP can have nested SPs.
SP cannot output 2D data (cannot return and output table variables).
SP cannot be called from a SELECT statement. It can be executed using only a EXEC/EXECUTE statement

38)What are the advantages of SP?

Answer)Precompiled code hence faster.

They allow modular programming, which means it allows you to break down a big chunk of code into smaller pieces of codes. This way the code will be more readable and more easier to manage.

Reusability.

Can enhance security of your application. Users can be granted permission to execute SP without having to have direct permissions on the objects referenced in the procedure.

Can reduce network traffic. An operation of hundreds of lines of code can be performed through single statement that executes the code in procedure rather than by sending hundreds of lines of code over the network.

SPs are pre-compiled, which means it has to have an Execution Plan so every time it gets executed after creating a new Execution Plan, it will save up to 70% of execution time. Without it, the SPs are just like any regular TSQL statements.

39)What is User Defined Functions (UDF)?

www.smartdatacamp.com

Answer)UDFs are a database object and a precompiled set of TSQL statements that can accept parameters, perform complex business calculation, and return of the action as a value.

The return value can either be single scalar value or result set-2D data. ○ UDFs are also precompiled and their execution plan is saved. PASSING INPUT PARAMETER(S) IS/ARE OPTIONAL, BUT MUST HAVE A RETURN STATEMENT.

40)What is the difference between Stored Procedure and UDF?

Answer)Stored Procedure:
may or may not return any value. When it does, it must be scalar INT. Can create temporary tables.
Can have robust error handling in SP (TRY/CATCH, transactions). Can include any DDL and DML statements.
UDF:
must return something, which can be either scalar/table valued. Cannot access to temporary tables.
No robust error handling available in UDF like TRY/ CATCH and transactions. Cannot have any DDL and can do DML only with table variables.

41)What are the types of UDF?

Answer)1. Scalar
Deterministic UDF: UDF in which particular input results in particular output. In other words, the output depends on the input.
Non-deterministic UDF: UDF in which the output does not directly depend on the input.
2. In-line UDF:
UDFs that do not have any function body(BEGIN...END) and has only a RETURN statement. In-line UDF must return 2D data.
3. Multi-line or Table Valued Functions:
It is an UDF that has its own function body (BEGIN ... END) and can have multiple SQL statements that return a single output. Also must return 2D data in the form of table variable

42)What is the difference between a nested UDF and recursive UDF?

Answer)Nested UDF: calling an UDF within an UDF
Recursive UDF: calling an UDF within itself

43)What is a Trigger?

Answer)It is a precompiled set of TSQL statements that are automatically executed on a particular DDL,DML or log-on event.
Triggers do not have any parameters or return statement.
Triggers are the only way to access to the INSERTED and DELETED tables (aka. Magic Tables).
You can DISABLE/ENABLE Triggers instead of DROPPING them:
DISABLE TRIGGER <name> ON <table/view name>/DATABASE/ALL SERVER
ENABLE TRIGGER <name> ON <table/view name>/DATABASE/ALL SERVER

44)What are the types of Triggers?

Answer)1. DML Trigger

DML Triggers are invoked when a DML statement such as INSERT, UPDATE, or DELETE occur which modify data in a specified TABLE or VIEW.

A DML trigger can query other tables and can include complex TSQL statements. They can cascade changes through related tables in the database.

They provide security against malicious or incorrect DML operations and enforce restrictions that are more complex than those defined with constraints.

2. DDL Trigger

Pretty much the same as DML Triggers but DDL Triggers are for DDL operations. DDL Triggers are at the database or server level (or scope).

DDL Trigger only has AFTER. It does not have INSTEAD OF.

3. Logon Trigger

Logon triggers fire in response to a logon event.

This event is raised when a user session is established with an instance of SQL server. Logon TRIGGER has server scope.

45)What are inserted and deleted tables (aka. magic tables)?

Answer)They are tables that you can communicate with between the external code and trigger body.

The structure of inserted and deleted magic tables depends upon the structure of the table in a DML statement.

UPDATE is a combination of INSERT and DELETE, so its old record will be in the deleted table and its new record will be stored in the inserted table.

46)What are some String functions to remember? LEN(string): returns the length of string.

Answer)UPPER(string) & LOWER(string): returns its upper/lower string

www.smartdatacamp.com

LTRIM(string) & RTRIM(string): remove empty string on either ends of the string LEFT(string):

extracts a certain number of characters from left side of the string RIGHT(string): extracts a certain

number of characters from right side of the string SUBSTRING(string, starting_position, length):

returns the sub string of the string REVERSE(string): returns the reverse string of the string

Concatenation: Just use + sign for it

REPLACE(string, string_replaced, string_replace_with)

47)What are the three different types of Error Handling?

Answer)1. TRY CATCH

The first error encountered in a TRY block will direct you to its CATCH block ignoring the rest of the code in the TRY block will generate an error or not.

2. @@error

stores the error code for the last executed SQL statement. If there is no error, then it is equal to 0. If there is an error, then it has another number (error code).

3. RAISERROR() function

A system defined function that is used to return messages back to applications using the same format which SQL uses for errors or warning message.

48)Explain about Cursors ..?

Answer)Cursors are a temporary database object which are used to loop through a table on row-by-row basis. There are five types of cursors:

1. Static: shows a static view of the data with only the changes done by session which opened the cursor.

2. Dynamic: shows data in its current state as the cursor moves from record-to-record.

3. Forward Only: move only record-by-record

4. Scrolling: moves anywhere.

5. Read Only: prevents data manipulation to cursor data set.

49)What is the difference between Table scan and seek ?

Answer)Scan: going through from the first page to the last page of an offset by offset or row by row.

Seek: going to the specific node and fetching the information needed.

Seek is the fastest way to find and fetch the data. So if you see your Execution Plan and if all of them is a seek, that means it's optimized.

50)Why are the DML operations are slower on Indexes?

Answer)It is because the sorting of indexes and the order of sorting has to be always maintained.

When inserting or deleting a value that is in the middle of the range of the index, everything has to be rearranged again. It cannot just insert a new value at the end of the index.

51)What is a heap (table on a heap)?

Answer)When there is a table that does not have a clustered index, that means the table is on a heap.

Ex:

Following table 'Emp' is a table on a heap.

SELECT * FROM Emp WHERE ID BETWEEN 2 AND 4 -- This will do scanning.

52)What is the architecture in terms of a hard disk, extents and pages?

Answer)A hard disk is divided into Extents.

Every extent has eight pages.

Every page is 8KBs (8060 bytes).

53)What are the nine different types of Indexes?

Answer)1. Clustered
2. Non-clustered
3. Covering
4. Full Text Index
5. Spatial
6. Unique
7. Filtered
8. XML
9. Index View

www.smartdatacamp.com

54)What is a Clustering Key?

Answer) It is a column on which I create any type of index is called a Clustering Key for that particular index

55)Explain about a Clustered Index.?

Answer)Unique Clustered Indexes are automatically created when a PK is created on a table.

But that does not mean that a column is a PK only because it has a Clustered Index.

Clustered Indexes store data in a contiguous manner. In other words, they cluster the data into a certain spot on a hard disk continuously The clustered data is ordered physically. You can only have one CI on a table.

56)What happens when Clustered Index is created?

Answer)First, a B-Tree of a CI will be created in the background.

Then it will physically pull the data from the heap memory and physically sort the data based on the clustering key.

Then it will store the data in the leaf nodes.

Now the data is stored in your hard disk in a continuous manner.

57)What are the four different types of searching information in a table?

Answer)1. Table Scan -> the worst way
2. Table Seek -> only theoretical, not possible
3. Index Scan -> scanning leaf nodes
4. Index Seek -> getting to the node needed, the best way

58)What is Fragmentation .?

Answer)Fragmentation is a phenomenon in which storage space is used inefficiently.

In SQL Server, Fragmentation occurs in case of DML statements on a table that has an index.

When any record is deleted from the table which has any index, it creates a memory bubble which causes fragmentation.

Fragmentation can also be caused due to page split, which is the way of building B-Tree dynamically according to the new records coming into the table.

Taking care of fragmentation levels and maintaining them is the major problem for Indexes.

Since Indexes slow down DML operations, we do not have a lot of indexes on OLTP, but it is recommended to have many different indexes in OLAP.

59)What are the two types of fragmentation?

Answer)1. Internal Fragmentation

It is the fragmentation in which leaf nodes of a B-Tree is not filled to its fullest capacity and contains memory bubbles.

2. External Fragmentation

It is fragmentation in which the logical ordering of the pages does not match the physical ordering of the pages on the secondary storage device.

60)What are Statistics?

Answer)Statistics allow the Query Optimizer to choose the optimal path in getting the data from the underlying table.

Statistics are histograms of max 200 sampled values from columns separated by intervals.
Every statistic holds the following info:
1. The number of rows and pages occupied by a table's data
2. The time that statistics was last updated
3. The average length of keys in a column
4. Histogram showing the distribution of data in column

61)What are some optimization techniques in SQL?

Answer)1. Build indexes. Using indexes on a table, It will dramatically increase the performance of your read operation because it will allow you to perform index scan or index seek depending on your search predicates and select predicates instead of table scan. Building non-clustered indexes, you could also increase the performance further.

2. You could also use an appropriate filtered index for your non clustered index because it could avoid performing a key lookup.

3. You could also use a filtered index for your non-clustered index since it allows you to create an index on a particular part of a table that is accessed more frequently than other parts.

4. You could also use an indexed view, which is a way to create one or more clustered indexes on the same table. In that way, the query optimizer will consider even the clustered keys on the indexed views so there might be a possible faster option to execute your query.

www.smartdatacamp.com

5. Do table partitioning. When a particular table as a billion of records, it would be practical to partition a table so that it can increase the read operation performance. Every partitioned table will be considered as physical smaller tables internally.

6. Update statistics for TSQL so that the query optimizer will choose the most optimal path in getting the data from the underlying table. Statistics are histograms of maximum 200 sample values from columns separated by intervals.

7. Use stored procedures because when you first execute a stored procedure, its execution plan is stored and the same execution plan will be used for the subsequent executions rather than generating an execution plan every time.

8. Use the 3 or 4 naming conventions. If you use the 2 naming convention, table name and column name, the SQL engine will take some time to find its schema. By specifying the schema name or even server name, you will be able to save some time for the SQL server.

9. Avoid using SELECT *. Because you are selecting everything, it will decrease the performance. Try to select columns you need.

10. Avoid using CURSOR because it is an object that goes over a table on a row-by-row basis, which is similar to the table scan. It is not really an effective way.

11. Avoid using unnecessary TRIGGER. If you have unnecessary triggers, they will be triggered needlessly. Not only slowing the performance down, it might mess up your whole program as well.

12. Manage Indexes using RECOMPILE or REBUILD. The internal fragmentation happens when there are a lot of data bubbles on the leaf nodes of the b-tree and the leaf nodes are not used to its fullest capacity. By recompiling, you can push the actual data on the b-tree to the left side of the leaf level and push the memory bubble to the right side. But it is still a temporary solution because the memory bubbles will still exist and won't be still accessed much. The external fragmentation occurs when the logical ordering of the b-tree pages does not match the physical ordering on the hard disk. By rebuilding, you can cluster them all together, which will solve not only the internal but also the external fragmentation issues. You can check the status of the fragmentation by using Data Management Function, sys.dm_db_index_physical_stats(db_id, table_id, index_id, partition_num, flag), and looking at the columns, avg_page_space_used_in_percent for the internal fragmentation and avg_fragmentation_in_percent for the external fragmentation.

13. Try to use JOIN instead of SET operators or SUB-QUERIES because set operators and subqueries are slower than joins and you can implement the features of sets and sub-queries using joins.

14. Avoid using LIKE operators, which is a string matching operator but it is mighty slow.

15. Avoid using blocking operations such as order by or derived columns.

16. For the last resort, use the SQL Server Profiler. It generates a trace file, which is a really detailed version of execution plan. Then DTA (Database Engine Tuning Advisor) will take a trace file as its input and analyzes it and gives you the recommendation on how to improve your query further.

62)How do you present the following tree in a form of a table?

Answer)
A
/ \
B C
/ \ / \
D E F G
CREATE TABLE tree (node CHAR(1), parent Node CHAR(1), [level] INT) INSERT INTO tree
VALUES ('A', null, 1),
('B', 'A', 2),
('C', 'A', 2),
('D', 'B', 3),
('E', 'B', 3),
('F', 'C', 3),
('G', 'C', 3)
SELECT * FROM tree
Result:
A NULL 1
B A 2
C A 2
D B 3
E B 3
F C 3
G C 3

63)How do you reverse a string without using REVERSE (string) ?

Answer)CREATE PROC rev (@string VARCHAR(50)) AS
BEGIN
DECLARE @new_string VARCHAR(50) = ''
DECLARE @len INT = LEN(@string)
WHILE (@len <> 0)
BEGIN
DECLARE @char CHAR(1) = SUBSTRING(@string, @len, 1) SET @new_string = @new_string +
@char
SET @len = @len - 1
END
PRINT @new_string
END
EXEC rev dinesh

64)What is Deadlock?

www.smartdatacamp.com

Answer)Deadlock is a situation where, say there are two transactions, the two transactions are waiting for each other to release their locks.

The SQL automatically picks which transaction should be killed, which becomes a deadlock victim, and roll back the change for it and throws an error message for it.

65)What is a Fact Table?

Answer)The primary table in a dimensional model where the numerical performance measurements (or facts) of the business are stored so they can be summarized to provide information about the history of the operation of an organization.

We use the term fact to represent a business measure. The level of granularity defines the grain of the fact table.

66)What is a Dimension Table?

Answer)Dimension tables are highly denormalized tables that contain the textual descriptions of the business and facts in their fact table.

Since it is not uncommon for a dimension table to have 50 to 100 attributes and dimension tables tend to be relatively shallow in terms of the number of rows, they are also called a wide table.

A dimension table has to have a surrogate key as its primary key and has to have a business/alternate key to link between the OLTP and OLAP.

67)What are the types of Measures?

Answer)Additive: measures that can be added across all dimensions (cost, sales).

Semi-Additive: measures that can be added across few dimensions and not with others.

Non-Additive: measures that cannot be added across all dimensions (stock rates).

68)What is a Star Schema?

Answer)It is a data warehouse design where all the dimensions tables in the warehouse are directly connected to the fact table.

The number of foreign keys in the fact table is equal to the number of dimensions.

It is a simple design and hence faster query.

69)What is a Snowflake Schema?

Answer)It is a data warehouse design where at least one or more multiple dimensions are further normalized.

Number of dimensions > number of fact table foreign keys

Normalization reduces redundancy so storage wise it is better but querying can be affected due to the excessive joins that need to be performed.

70)What is granularity?

Answer)The lowest level of information that is stored in the fact table. ○ Usually determined by the time dimension table.

The best granularity level would be per transaction but it would require a lot of memory.

71)What is a Surrogate Key?

Answer)It is a system generated key that is an identity column with the initial value and incremental value and ensures the uniqueness of the data in the dimension table.

Every dimension table must have a surrogate key to identify each record!!!

72)What are some advantages of using the Surrogate Key in a Data Warehouse?

Answer)1. Using a SK, you can separate the Data Warehouse and the OLTP: to integrate data coming from heterogeneous sources, we need to differentiate between similar business keys from the OLTP. The keys in OLTP are the alternate key (business key).

2. Performance: The fact table will have a composite key. If surrogate keys are used, then in the fact table, we will have integers for its foreign keys. This requires less storage than VARCHAR. The queries will run faster when you join on integers rather than VARCHAR. The partitioning done on SK will be faster as these are in sequence.

3. Historical Preservation: A data warehouse acts as a repository of historical data so there will be various versions of the same record and in order to differentiate between them, we need a SK then we can keep the history of data.

4. Special Situations (Late Arriving Dimension): Fact table has a record that doesn't have a match yet in the dimension table. Surrogate key usage enables the use of such a 'not found' record as a SK is not dependent on the ETL process.

73)What is the datatype difference between a fact and dimension tables?

Answer)1. Fact Tables
They hold numeric data.

www.smartdatacamp.com

They contain measures.
They are deep.
2. Dimensional Tables
They hold textual data.
They contain attributes of their fact tables.
They are wide.

74)What are the types of dimension tables?

Answer)1. Conformed Dimensions when a particular dimension is connected to one or more fact tables. ex) time dimension

2. Parent-child Dimensions A parent-child dimension is distinguished by the fact that it contains a hierarchy based on a recursive relationship. when a particular dimension points to its own surrogate key to show an unary relationship.

3. Role Playing Dimensions when a particular dimension plays different roles in the same fact table. ex) dim_time and orderDateKey, shippedDateKey...usually a time dimension table. Role-playing dimensions conserve storage space, save processing time, and improve database manageability .

4. Slowly Changing Dimensions: A dimension table that have data that changes slowly that occur by inserting and updating of records.

1. Type 0: columns where changes are not allowed - no change ex) DOB, SSNm

2. Type 1: columns where its values can be replaced without adding its new row - replacement

3. Type 2: for any change for the value in a column, a new record it will be added - historical data.

Previous values are saved in records marked as outdated. For even a single type 2 column, startDate, EndDate, and status are needed.

4. Type 3: advanced version of type 2 where you can set up the upper limit of history which drops the oldest record when the limit has been reached with the help of outside SQL implementation. Type 0 ~ 2 are implemented on the column level.

5. Degenerated Dimensions: a particular dimension that has an one-to-one relationship between itself and the fact table. When a particular Dimension table grows at the same rate as a fact table, the actual dimension can be removed and the dimensions from the dimension table can be inserted into the actual fact table.

You can see this mostly when the granularity level of the the facts are per transaction. E.g. The dimension salesorderdate (or other dimensions in DimSalesOrder would grow everytime a sale is made therefore the dimension (attributes) would be moved into the fact table.

6. Junk Dimensions: holds all miscellaneous attributes that may or may not necessarily belong to any other dimensions. It could be yes/no, flags, or long open-ended text data.

75)What is your strategy for the incremental load?

Answer)The combination of different techniques for the incremental load in my previous projects; timestamps, CDC (Change Data Capture), MERGE statement and CHECKSUM() in TSQL, LEFT OUTER JOIN, TRIGGER, the Lookup Transformation in SSIS

76)What is CDC?

Answer)CDC (Change Data Capture) is a method to capture data changes, such as INSERT, UPDATE and DELETE, happening in a source table by reading transaction log files. Using CDC in the process of an incremental load, you are going to be able to store the changes in a SQL table, enabling us to apply the changes to a target table incrementally.

In data warehousing, CDC is used for propagating changes in the source system into your data warehouse, updating dimensions in a data mart, propagating standing data changes into your data warehouse and such.
The advantages of CDC are:
- It is almost real time ETL.
- It can handle small volume of data.
- It can be more efficient than replication.
- It can be auditable.
- It can be used to configurable clean up.
Disadvantages of CDC are:
- Lots of change tables and functions
- Bad for big changes e.g. truncate & reload Optimization of CDC:
- Stop the capture job during load
- When applying changes to target, it is ideal to use merge.

77)What is the difference between a connection and session?

Answer)Connection: It is the number of instance connected to the database. An instance is modelized soon as the application is open again.

Session: A session run queries.In one connection, it allowed multiple sessions for one connection.

78)What are all different types of collation sensitivity?

Answer)Following are different types of collation sensitivity -
Case Sensitivity - A and a and B and b.
Accent Sensitivity.
Kana Sensitivity - Japanese Kana characters.
Width Sensitivity - Single byte character and double byte character.

79)What is CLAUSE?

www.smartdatacamp.com

Answer)SQL clause is defined to limit the result set by providing condition to the query. This usually filters

some rows from the whole set of records.

Example - Query that has WHERE condition Query that has HAVING condition.

80)What is Union, minus and Interact commands?

Answer)UNION operator is used to combine the results of two tables, and it eliminates duplicate rows from the tables.

MINUS operator is used to return rows from the first query but not from the second query. Matching records of first and second query and other rows from the first query will be displayed as a result set.

INTERSECT operator is used to return rows returned by both the queries.

81)How to fetch common records from two tables?

Answer)Common records result set can be achieved by -.

Select studentID from student. INTERSECT Select StudentID from Exam

82)How to fetch alternate records from a table?

Answer)Records can be fetched for both Odd and Even row numbers -.
To display even numbers-.
Select studentId from (Select rowno, studentId from student) where mod(rowno,2)=0 To display odd
numbers-.
Select studentId from (Select rowno, studentId from student) where mod(rowno,2)=1 from (Select rowno, studentId from student) where mod(rowno,2)=1.[/sql]

83)How to select unique records from a table?

Answer)Select unique records from a table by using DISTINCT keyword.

Select DISTINCT StudentID, StudentName from Student.

84)How to remove duplicate rows from table?

Answer)Step 1: Selecting Duplicate rows from table
Select rollno FROM Student WHERE ROWID <>
(Select max (rowid) from Student b where rollno=b.rollno);
Step 2: Delete duplicate rows
Delete FROM Student WHERE ROWID <>
(Select max (rowid) from Student b where rollno=b.rollno);

85)What is ROWID and ROWNUM in SQL?

Answer)RowID
1.ROWID is nothing but Physical memory allocation
2.ROWID is permanant to that row which identifies the address of that row.
3.ROWID is 16 digit Hexadecimal number which is uniquely identifies the rows.
4.ROWID returns PHYSICAL ADDRESS of that row.
5. ROWID is automatically generated unique id of a row and it is generated at the time of insertion of row.
6. ROWID is the fastest means of accessing data.
ROWNUM:
1. ROWNUM is nothing but the sequence which is allocated to that data retreival bunch.
2. ROWNUM is tempararily allocated sequence to the rows.
3.ROWNUM is numeric sequence number allocated to that row temporarily.
4.ROWNUM returns the sequence number to that row.
5. ROWNUM is an dynamic value automatically retrieved along with select statement output.
6.ROWNUM is not related to access of data

86)How to find count of duplicate rows?

Answer)Select rollno, count (rollno) from Student

Group by rollno Having count (rollno)>1 Order by count (rollno) desc;

87)How to find Third highest salary in Employee table using self-join?

Answer)Select * from Employee a Where 3 = (Select Count (distinct Salary) from Employee where a.salary<=b.salary;

88)How to display following using query?

Answer)
*
**
***

www.smartdatacamp.com

We cannot use dual table to display output given above. To display output use any table. I am using Student table.
SELECT lpad ('*', ROWNUM,'*') FROM Student WHERE ROWNUM <4;

89)How to display Date in DD-MON-YYYY table?

Answer)Select to_date (Hire_date,'DD-MON-YYYY') Date_Format from Employee;

90)If marks column contain the comma separated values from Student table. How to calculate the count of that comma separated values?

Answer)Student Name Marks
Dinesh 30,130,20,4
Kumar 100,20,30
Sonali 140,10
Select Student_name, regexp_count (marks,',') + As "Marks Count" from Student;

91)What is query to fetch last day of previous month in oracle?

Answer)Select LAST_DAY (ADD_MONTHS (SYSDATE,-1)) from dual;

92)How to display the String vertically in Oracle?

Answer)SELECT SUBSTR ('AMIET', LEVEL, 1) FROM dual Connect by level <= length ('AMIET');

93)How to display departmentwise and monthwise maximum salary?

Answer)Select Department_no, TO_CHAR (Hire_date,'Mon') as Month from Employee group by Department_no, TO_CHAR (Hire_date,'mon');

94)How to calculate number of rows in table without using count function?

Answer)Select table_name, num_rows from user_tables where table_name='Employee';

Tip: User needs to use the system tables for the same. So using user_tables user will get the number of rows in the table

95)How to fetch common records from two different tables which has not any joining condition?

Answer)Select * from Table1

Intersect

Select * from Table2;

96)Explain Execution Plan.?

Answer)Query optimizer is a part of SQL server that models the way in which the relational DB engine works and comes up with the most optimal way to execute a query.

Query Optimizer takes into account amount of resources used, I/O and CPU processing time etc. to generate a plan that will allow query to execute in most efficient and faster manner. This is known as EXECUTION PLAN.

Optimizer evaluates a number of plans available before choosing the best and faster on available. Every query has an execution plan.

Definition by the mod: Execution Plan is a plan to execute a query with the most optimal way which is generated by Query Optimizer.

Query Optimizer analyzes statistics, resources used, I/O and CPU processing time and etc. and comes up with a number of plans. Then it evaluates those plans and the most optimized plan out of the plans is Execution Plan. It is shown to users as a graphical flow chart that should be read from right to left and top to bottom.

www.smartdatacamp.com

Scala

Scala combines object-oriented and functional programming in one concise, high-level language. Scala's static types help avoid bugs in complex applications, and its JVM and JavaScript runtimes let you build high-performance systems with easy access to huge ecosystems of libraries.

Scala is a strong statically typed general-purpose programming language which supports both object-oriented programming and functional programming. Designed to be concise, many of Scala's design decisions are aimed to address criticisms of Java.

1) What is Scala?

Answer)Scala is a general-purpose programming language providing support for both functional and ObjectOriented programming.

2) What is tail-recursion in Scala?

Answer)There are several situations where programmers have to write functions that are recursive in nature.

The main problem with recursive functions is that, it may eat up all the allocated stack space.

To overcome this situation, Scala compiler provides a mechanism “tail recursion” to optimize these recursive functions so that it does not create new stack space, instead uses the current function stack space.

To qualify for this, annotation “@annotation.tailrec” has to be used before defining the function and recursive call has to be the last statement, then only the function will compile otherwise, it will give an error.

3)What are traits in Scala?

Answer)Traits are used to define object types specified by the signature of the supported methods.

Scala allows to be partially implemented but traits may not have constructor parameters. A trait consists of method and field definition, by mixing them into classes it can be reused.

4)Who is the father of Scala programming language?

Answer)Martin Oderskey, a German computer scientist, is the father of Scala programming language.

5)What are case classes in Scala?

Answer)Case classes are standard classes declared with a special modifier case. Case classes export their constructor parameters and provide a recursive decomposition mechanism through pattern matching.

The constructor parameters of case classes are treated as public values and can be accessed directly. For a case class, companion objects and its associated method also get generated automatically. All the methods in the class, as well, methods in the companion objects are generated based on the parameter list. The only advantage of Case class is that it automatically generates the methods from the parameter list.

6)What is the super class of all classes in Scala?

Answer)In Java, the super class of all classes (Java API Classes or User Defined Classes) is java.lang.Object.

In the same way in Scala, the super class of all classes or traits is “Any” class.

Any class is defined in scala package like “scala.Any”

7)What is a ‘Scala Set’? What are methods through which operation sets are expressed?

Answer)Scala set is a collection of pairwise elements of the same type. Scala set does not contain any duplicate elements. There are two kinds of sets, mutable and immutable.

8)What is a Scala Map?

Answer)Scala Map is a collection of key value pairs wherein the value in a map can be retrieved using the key.

Values in a Scala Map are not unique, but the keys are unique. Scala supports two kinds of mapsmutable and immutable. By default, Scala supports immutable map and to make use of the mutable map, programmers must import the scala.collection.mutable.Map class explicitly.

When programmers want to use mutable and immutable map together in the same program then the mutable map can be accessed as mutable.map and the immutable map can just be accessed with the name of the map

www.smartdatacamp.com

9)Name two significant differences between a trait and an abstract class.

Answer)Abstract classes have constructors with zero or more parameters while traits do not; a class can extend any number of traits but only one abstract class

10)What is the use of tuples in Scala?

Answer)Scala tuples combine a fixed number of items together so that they can be passed around as whole. A tuple is immutable and can hold objects with different types, unlike an array or list.

11)What do you understand by a closure in Scala?

Answer)A closure is also known as an anonymous function whose return value depends upon the value of the variables declared outside the function.

12)What do you understand by Implicit Parameter?

Answer)Wherever, we require that function could be invoked without passing all the parameters, we use implicit parameter.

We provide the default values for all the parameters or parameters which we want to be used as implicit.

When the function is invoked without passing the implicit parameters, local value of that parameter is used.

We need to use implicit keyword to make a value, function parameter or variable as implicit.

13)What is the companion object in Scala?

Answer)A companion object is an object with the same name as a class or trait and is defined in the same source file as the associated file or trait.

A companion object differs from other objects as it has access rights to the class/trait that other objects do not.

In particular it can access methods and fields that are private in the class/trait.

14)What are the advantages of Scala Language?

Answer)Advantages of Scala Language:-

- Simple and Concise Code
- Very Expressive Code
- More Readable Code
- 100% Type-Safe Language
- Immutability and No Side-Effects
- More Reusable Code
- More Modularity
- Do More with Less Code
- Supports all OOP Features
- Supports all FP Features. Highly Functional.
- Less Error Prone Code
- Better Parallel and Concurrency Programming
- Highly Scalable and Maintainable code
- Highly Productivity
- Distributed Applications
- Full Java Interoperability
- Powerful Scala DSLs available

15)What are the major drawbacks of Scala Language?

Answer)Drawbacks of Scala Language:-

- Less Readable Code
- Bit tough to Understand the Code for beginners
- Complex Syntax to learn
- Less Backward Compatibility

16)What is Akka, Play, and Sleek in Scala?

Answer)Akka is a concurrency framework in Scala which uses Actor based model for building highly concurrent, distributed, and resilient message-driven applications on the JVM.

It uses high-level abstractions like Actor, Future, and Stream to simplify coding for concurrent applications. It also provides load balancing, routing, partitioning, and adaptive cluster management.

www.smartdatacamp.com

If you are interested in learning Akka,

17)What is 'Unit' and '()' in Scala?

Answer)The 'Unit' is a type like void in Java. You can say it is a Scala equivalent of the void in Java, while still providing the language with an abstraction over the Java platform. The empty tuple '()' is a term representing a Unit value in Scala.

18)What is the difference between a normal class and a case class in Scala?

Answer)Following are some key differences between a case class and a normal class in Scala:

- case class allows pattern matching on it.

- you can create instances of case class without using the new keyword

- equals(), hashcode() and toString() method are automatically generated for case classes in Scala

- Scala automatically generate accessor methods for all constructor argument

19)What are High Order Functions in Scala?

Answer)High order functions are functions that can receive or return other functions. Common examples in Scala are the filter, map, and flatMap functions, which receive other functions as arguments

20)Which Scala library is used for functional programming?

Answer)Scala library has purely functional data structures that complement the standard Scala library. It has pre-defined set of foundational type classes like Monad, Functor, etc.

21)What is the best scala style checker tool available for play and scala based applications?

Answer)Scalastyle is best Scala style checker tool available for Play and Scala based applications.

Scalastyle observes the Scala source code and indicates potential problems with it.

It has three separate plug-ins to supports the following build tools:

SBT

Maven

Gradle

22)What is the difference between concurrency and parallelism?

Answer)When several computations execute sequentially during overlapping time periods it is referred to as concurrency whereas when processes are executed simultaneously it is known as parallelism.

Parallel collection, Futures and Async library are examples of achieving parallelism in Scala.

23)What is the difference between a Java method and a Scala function?

Answer)Scala function can be treated as a value. It can be assigned to a val or var, or even returned from another function, which is not possible in Java.

Though Java 8 brings lambda expression which also makes function as a first-class object, which means you can pass a function to a method just like you pass an object as an argument.

See here to learn more about the difference between Scala and Java.

24)What is the difference between Function and Method in Scala?

Answer)Scala supports both functions and methods. We use same syntax to define functions and methods, there is no syntax difference.

However, they have one minor difference:

We can define a method in a Scala class or trait.

Method is associated with an object (An instance of a Class).

We can call a method by using an instance of a Class.

We cannot use a Scala Method directly without using object.

Function is not associated with a class or trait.

It is defined in a Scala Package.

We can access functions without using objects, like Java's Static Methods

25)What is Extractor in Scala?

Answer)In Scala, Extractor is used to decompose or disassemble an object into its parameters (or components)

26)Is Scala a Pure OOP Language?

www.smartdatacamp.com

Answer)Yes, Scala is a Pure Object-Oriented Programming Language because in Scala, everything is an Object, and everything is a value.

Functions are values and values are Objects. Scala does not have primitive data types and does not have static members.

27)Is Java a pure OOP Language?

Answer)Java is not a Pure Object-Oriented Programming (OOP) Language because it supports the following two Non-OOP concepts:

Java supports primitive data types.

They are not objects.

Java supports Static members.

They are not related to objects.

28)Does Scala support Operator Overloading? Scala supports Operator Overloading.

Answer)Scala has given this flexibility to Developer to decide which methods/functions name should use.

When we call 4 + 5 that means '+' is not an operator, it is a method available in Int class (or it's implicit type).

Internally, this call is converted into "4.+(5)".

29)Does Java support Operator Overloading?

Answer)Java does not support Operator Overloading.

30)What are the default imports in Scala Language?

Answer)We know, java.lang is the default package imported into all Java Programs by JVM automatically.

We don't need to import this package explicitly.

In the same way, the following are the default imports available in all Scala Programs:

java.lang package

Scala package

scala.PreDef

31)What is an Expression?

Answer)Expression is a value that means it will evaluate to a Value. As an Expression returns a value, we can assign it to a variable.

Example:- Scalas If condition, Javas Ternary operator

32)What is a Statement? Difference between Expression and Statement?

Answer)Statement defines one or more actions or operations.

That means Statement performs actions.

As it does not return a value, we cannot assign it to a Variable.

Example:- Java's If condition.

33)What is the difference between Java's "If...Else" and Scala's "If..Else"?

Answer)Java's "If..Else":

In Java, "If..Else" is a statement, not an expression. It does not return a value and cannot assign it to a variable.

Example:-
int year;
if(count == 0)
year = 2018;
else
year = 2017;
Scala's "If..Else":
In Scala, "If..Else" is an expression. It evaluates a value i.e. returns a value. We can assign it to a variable.
val year = if(count == 0) 2018 else 2017
NOTE:-Scala's "If..Else" works like Java's Ternary Operator. We can use Scala's "If..Else" like Java's "If..Else" statement as shown below:
val year = 0
if(count == 0)
year = 2018
else
year = 2017

34)How to compile and run a Scala program?

www.smartdatacamp.com

Answer)You can use Scala compiler scalac to compile Scala program (like javac) and scala command to run them (like scala)

35)How to tell Scala to look into a class file for some Java class?

Answer)We can use -classpath argument to include a JAR in Scala's classpath, as shown below

$ scala -classpath jar

Alternatively, you can also use CLASSPATH environment variable.

36)What is the difference between a call-by-value and call-by-name parameter?

Answer)The main difference between a call-by-value and a call-by-name parameter is that the former is computed before calling the function, and the latter is evaluated when accessed.

37)What exactly is wrong with a recursive function that is not tail-recursive?

Answer)You run the risk of running out of stack space and thus throwing an exception.

38)What is the difference between var and value?

Answer)In scala, you can define a variable using either a, val or var keywords.

The difference between val and var is, var is much like java declaration, but valis little different.

We cannot change the reference to point to another reference, once the variable is declared using val.

The variable defined using var keywords are mutable and can be changed any number of times.

39)What is scala anonymous function?

Answer)In a source code, anonymous functions are called 'function literals' and at run time, function literals are instantiated into objects called function values.

Scala provides a relatively easy syntax for defining anonymous functions.

40)What is function currying in scala?

Answer)Currying is the technique of transforming a function that takes multiple arguments into a function that takes a single argument Many of the same techniques as language like Haskell and LISP are supported by Scala. Function currying is one of the least used and misunderstood one.

41)What do you understand by “Unit” and “()” in Scala?

Answer)Unit is a subtype of scala.anyval and is nothing but Scala equivalent of Java void that provides the Scala with an abstraction of the java platform. Empty tuple i.e. () in Scala is a term that represents unit value.

42)What’s the difference ‘Nil’, ‘Null’, ‘None’ and ‘Nothing’ in Scala?

Answer)Null - It’s a sub-type of AnyRef type in Scala Types hierarchy. As Scala runs on JVM, it uses NULL
to provide the compatibility with Java null keyword, or in Scala terms, to provide type for null keyword, Null type exists. It represents the absence of type information for complex types that are
inherited from AnyRef.
Nothing - It’s a sub-type of all the types exists in Scala Types hierarchy. It helps in providing the return type for the operations that can affect a normal program’s flow. It can only be used as a type, as
instantiation of nothing cannot be done. It incorporates all types under AnyRef and AnyVal. Nothing
is usually used as a return type for methods that have abnormal termination and result in an exception.
Nil - It’s a handy way of initializing an empty list since, Nil, is an object, which extends List [Nothing].
None - In programming, there are many circumstances, where we
unexpectedly received null for the methods we call. In java these are
handled using try/catch or left unattended causing errors in the program.
Scala provides a very graceful way of handling those situations. In cases,
where you don’t know, if you would be able to return a value as
expected, we can use Option [T]. It is an abstract class, with just two
sub-classes, Some [T] and none. With this, we can tell users that, the method might return a T of type
Some [T] or it might return none.

43)What is Lazy Evaluation?

Answer)Lazy Evaluation means evaluating program at run-time on-demand that means when clients access the program then only its evaluated.

The difference between “val” and “lazy val” is that “val” is used to define variables which are evaluated eagerly and “lazy val” is also used to define variables but they are evaluated lazily.

44)What is call-by-name?

Answer)Call-by-name means evaluates method/function parameters only when we need them, or we access them. If we don’t use them, then it does not evaluate them.

45)Does Scala and Java support call-by-name?

Answer)Scala supports both call-by-value and call-by-name function parameters. However, Java supports only call-by-value, but not call-by-name.

46)What is the difference between call-by-value and call-by-name function parameters?

Answer)Difference between call-by-value and call-by-name:

The major difference between these two are described below:

In Call-by-name, the function parameters are evaluated only whenever they are needed but not when the function is called.

In Call-by-value, the function parameters are evaluated when the function is called.

In Call-byvalue, the parameters are evaluated before executing function and they are evaluated only once irrespective of how many times we used them in that function.

In Call-by-name, the parameters are evaluated whenever we access them, and they are evaluated each time we use them in that function

47)What do you understand by apply and unapply methods in Scala?

Answer)Apply and unapply methods in Scala are used for mapping and unmapping data between form and model data. Apply method - Used to assemble an object from its components. For example, if we want to create an Employee object then use the two components firstName and lastName and compose the Employee object using the apply method. Unapply method - Used to decompose an object from its components. It follows the reverse process of apply method. So, if you have an employee object, it can be decomposed into two componentsfirstName and lastName.

48)What is an anonymous function in Scala?

Answer)Anonymous Function is also a Function, but it does not have any function name. It is also known as a Function Literal

49)What are the advantages of Anonymous Function/Function Literal in Scala?

Answer)The advantages of Anonymous Function/Function Literal in Scala:

We can assign a Function Literal to variable

We can pass a Function Literal to another function/method

We can return a Function Literal as another function/method result/return value.

50)What is the difference between unapply and apply, when would you use them?

Answer)Unapply is a method that needs to be implemented by an object in order for it to be an extractor.

Extractors are used in pattern matching to access an object constructor parameter. It's the opposite of a constructor.

The apply method is a special method that allows you to write someObject(params) instead of someObject.apply(params).

This usage is common in case classes, which contain a companion object with the apply method that allows the nice syntax to instantiate a new object without the new keyword.

51)What is the difference between a trait and an abstract class in Scala?

Answer)Here are some key differences between a trait and an abstract class in Scala:

A class can inherit from multiple traits but only one abstract class.

Abstract classes can have constructor parameters as well as type parameters.

Traits can have only type parameters.

For example, you can't say trait t(i: Int) {}; the iparameter is illegal.

Abstract classes are fully interoperable with Java. You can call them from Java code without any wrappers.

On the other hand, Traits are fully interoperable only if they do not contain any implementation code. See here to learn more about Abstract class in Java and OOP.

52)Can a companion object in Scala access the private members of its companion class in Scala?

www.smartdatacamp.com

Answer)According to the private access specifier, private members can be accessed only within that class, but Scala's companion object and class provide special access to private members.

A companion object can access all the private members of a companion class. Similarly, a companion class can access all the private members of companion objects.

53)What are scala variables?

Answer)Values and variables are two shapes that come in Scala. A value variable is constant and cannot be changed once assigned. It is immutable, while a regular variable, on the other hand, is mutable, and you can change the value. The two types of variables are var myVar : Int=0;

val myVal: Int=1;

54)Mention the difference between an object and a class ?

Answer)A class is a definition for a description. It defines a type in terms of methods and composition of other types.

A class is a blueprint of the object. While, an object is a singleton, an instance of a class which is unique.

An anonymous class is created for every object in the code, it inherits from whatever classes you declared object to implement.

55)What is the difference between val and var in Scala?

Answer)The val keyword stands for value and var stands for variable. You can use keyword val to store values, these are immutable, and cannot change once assigned. On the other hand, keyword var is used to create variables, which are values that can change after being set. If you try to modify a val, the compiler will throw an error. It is like the final variable in Java or const in C++.

56)What is the difference between Array and List in Scala?

Answer)Arrays are always Mutable whereas List is always Immutable.Once created, we can change Array values where as we cannot change List Object.Arrays are fixed-size data structures whereas List is variable-sized data structures. List's size is automatically increased or decreased based on its operations we perform on it.Arrays are Invariants whereas Lists are Covariant.

57)What is Type Inference in Scala?

Answer)Types can be inferred by the Scala Compiler at compile-time. It is known as Type Inference. Types means Data type or Result type. We use Types at many places in Scala programs like Variable types, Object types, Method/Function Parameter types, Method/Function return types etc. In simple words, determining the type of a variable or expression or object etc. at compile-time by compiler is known as “Type Inference”.

58)What is Eager Evaluation?

Answer)Eager Evaluation means evaluating program at compile-time or program deployment-time irrespective of clients are using that program or not.

59)What is guard in Scala for-Comprehension construct?

Answer)In Scala, for-comprehension construct has an if clause which is used to write a condition to filter some elements and generate new collection. This if clause is also known as “Guard”.If that guard is true, then add that element to new collection. Otherwise, it does not add that element to original collection 60. Why scala prefers immutability?Scala prefers immutability in design and in many cases uses it as default. Immutability can help when dealing with equality issues or concurrent programs.

60)What are the considerations you need to have when using Scala streams?

Answer)Streams in Scala are a type of lazy collection, which are created using starting element and then recursively generated using those elements. Streams are like a List, except that, elements are added only when they are accessed, hence “lazy”. Since streams are lazy in terms of adding elements, they can be unbounded also, and once the elements are added, they are cached. Since Streams can be unbounded, and all the values are computed at the time of access, programmers need to be careful on using methods which are not transformers, asit may result in java.lang.OutOfMemoryErrors. stream.max stream.size stream.sum

61)Differentiate between Array and List in Scala.

Answer)List is an immutable recursive data structure whilst array is a sequential mutable data structure.

Lists are covariant whilst array are invariants.The size of a list automatically increases or decreases based on the operations that are performed on it i.e. a list in Scala is a variable-sized data structure whilst an array is fixed size data structure.

www.smartdatacamp.com

62)Which keyword is used to define a function in Scala?

Answer)A function is defined in Scala using the def keyword. This may sound familiar to Python developers as Python also uses def to define a function.

63)What is Monad in Scala?

Answer)A monad is an object that wraps another object in Scala. It helps to perform the data manipulation of the underlying object, instead of manipulating the object directly

64)Is Scala statically-typed language?

Answer)Yes, Scala is a statically-typed language.

65)What is Statically Typed Language and What is Dynamically-Typed Language?

Answer)Statically Typed Language means that Type checking is done at compiletime by compiler, not at run time. Dynamically-Typed Language means that Type checking is done at run-time, not at compile time by compiler.

66)What is the difference between unapply and apply, when would you use them?

Answer)unapply is a method that needs to be implemented by an object in order for it to be an extractor.

Extractors are used in pattern matching to access an object constructor parameter. It's the opposite of a constructor.

The apply method is a special method that allows you to write someObject(params) instead of someObject.apply(params).

This usage is common in case classes, which contain a companion object with the apply method that allows the nice syntax to instantiate a new object without the new keyword.

67)What is Unit in Scala?

Answer)In Scala, Unit is used to represent No value or No Useful value. Unit is a final class defined in scala package that is scala.Unit.

68)What is the difference between Javas void and Scalas Unit?

Answer)Unit is something like Java's void. But they have few differences. Java's void does not any value. It is nothing.

Scalas Unit has one value ()

() is the one and only value of type Unit in Scala. However, there are no values of type void in Java.

Javas void is a keyword. Scalas Unit is a final class.Both are used to represent a method or function is not returning anything.

69)What is App in Scala?

Answer)In Scala, App is a trait defined in scala package like scala.App. It defines main method. If an Object or a Class extends this trait, then they will become as Scala Executable programs automatically because they will inherit main method from Application.

70)What is the use of Scala App?

Answer)The main advantage of using App is that we don't need to write main method. The main drawback of using App is that we should use same name "args" to refer command line argument because scala.App's main() method uses this name.

71)What are option, some and none in scala?

Answer)Option is a Scala generic type that can either be some generic value or none. Queue often uses it to represent primitives that may be null.

72)What is Scala Future?

Answer)Scala Future is a monadic collection, which starts a background task.

It is an object which holds the potential value or future value, which would be available after the task is completed.

It also provides various operations to further chain the operations or to extract the value.

Future also provide various call-back functions like onComplete, OnFailure, onSuccess to name a few, which makes Future a complete concurrent task class.

73)How it differs from java's Future class?

www.smartdatacamp.com

Answer)The main and foremost difference between Scalas Future and Javas Future class is that the later does not provide promises or callbacks operations. The only way to retrieve the result is Future.get () in Java.

74)What do you understand by diamond problem and how does Scala resolve this?

Answer)Multiple inheritance problem is referred to as the Deadly diamond problem or diamond problem.

The inability to decide on which implementation of the method to choose is referred to as the Diamond Problem in Scala.

Suppose say classes B and C both inherit from class A, while class D inherits from both class B and C.

Now while implementing multiple inheritance if B and C override some method from class A, there is a confusion and dilemma always on which implementation D should inherit.

This is what is referred to as diamond problem. Scala resolves diamond problem through the concept of Traits and class linearization rules.

75)What is the difference between == in Java and Scala?

Answer)Scala has more intuitive notion of equality. The == operator will automatically run the instance's equals method, rather than doing Java style comparison to check that two objects are the same reference. By the way, you can still check for referential equality by using eq method. In short, Java == operator compare references while Scala calls the equals() method. You can also read the difference between == and equals() in Java to learn more about how they behave in Java.

76)What is REPL in Scala? What is the use of Scala's REPL?

Answer)REPL stands for Read-Evaluate-Print Loop. We can pronounce it as ripple.

In Scala, REPL is acts as an Interpreter to execute Scala code from command prompt.

Thats why REPL is also known as Scala CLI(Command Line Interface) or Scala command-line shell.

The main purpose of REPL is that to develop and test small snippets of Scala code for practice purpose.

It is very useful for Scala Beginners to practice basic programs.

77)What are the similarities between Scalas Int and Javas java.lang.Integer?

Answer)Similarities between Scalas Int and Javas java.lang.Integer are Both are classes.Both are used to represent integer numbers. Both are 32-bit signed integers.

78)What are the differences between Scalas Int and Javas java.lang.Integer?

Answer)Differences between Scalas Int and Javas java.lang.Integer are Scala Int class does not implement Comparable interface. Javas java.lang.Integer class implements Comparable interface.

79)What is the relationship between Int and RichInt in Scala?

Answer)Javas Integer is something like Scalas Int and RichInt. RichInt is a final class defined in scala.runtime package like “scala.runtime.RichInt”.

In Scala, the Relationship between Int and RichInt is that when we use Int in a Scala program, it will automatically convert into RichInt to utilize all methods available in that Class. We can say that RichInt is an Implicit class of Int.

80)What is the best framework to generate rest api documentation for scala-based applications?

Answer)Swagger is the best tool for this purpose. It is very simple and open-source tool for generating REST APIs documentation with JSON for Scala-based applications.

If you use Play with Scala to develop your REST API, then use playswagger module for REST API documentation.

If you use Spray with Scala to develop your REST API, then use sprayswagger module for REST API documentation.

81)What is the use of Auxiliary Constructors in Scala?

Answer)Auxiliary Constructor is the secondary constructor in Scala declared using the keywords this and def.

The main purpose of using auxiliary constructors is to overload constructors. Just like in Java, we can provide implementation for different kinds of constructors so that the right one is invoked based on the requirements. Every auxiliary constructor in Scala should differ in the number of parameters or in data types.

82)How does yield work in Scala?

www.smartdatacamp.com

Answer)The yield keyword if specified before the expression, the value returned from every expression, will be returned as the collection.

The yield keyword is very useful, when there is a need, you want to use the return value of expression. The collection returned can be used the normal collection and iterate over in another loop.

83)What are the different types of Scala identifiers? There four types of Scala identifiers

Answer)Alpha numeric identifiers

Operator identifiers

Mixed identifiers

Literal identifiers

84)What are the different types of Scala literals?

Answer)The different types of literals in scala are

Integer literals

Floating point literals Boolean literals

Symbol literals

Character literals

String literals

Multi-Line strings

85)What is SBT? What is the best build tool to develop play and scala applications?

Answer)SBT stands for Scala Build Tool. Its a Simple Build Tool to develop Scalabased applications.

Most of the people uses SBT Build tool for Play and Scala Applications.

For example, IntelliJ IDEA Scala Plugin by default uses SBT as Build tool for this purpose

86)What is the difference between :: and ::: in Scala?

Answer):: and ::: are methods available in List class.

:: method is used to append an element to the beginning of the list.

And ::: method is used to concatenate the elements of a given list in front of this list.

:: method works as a cons operator for List class. Here 'cons' stands for construct.

::: method works as a concatenation operator for List class.

87)What is the difference between #:: and #::: in Scala?

Answer)#:: and #::: are methods available in Stream class

#:: method words as a cons operator for Stream class. Here 'cons' stands for construct.

#:: method is used to append a given element at beginning of the stream.

#::: method is used to concatenate a given stream at beginning of the stream.

88)What is the use of ??? in Scala-based Applications?

Answer)This ??? three question marks is not an operator, a method in Scala. It is used to mark a method which is In Progress that means Developer should provide implementation for that one.

89)What is the best Scala style checker tool available for Play and Scala based applications?

Answer)Scalastyle is best Scala style checker tool available for Play and Scala based applications.

Scalastyle observes our Scala source code and indicates potential problems with it. It has three separate plug-ins to supports the following build tools:

SBT

Maven

Gradle

It has two separate plug-ins to supports the following two IDEs:

IntelliJ IDEA

Eclipse IDE

90)How Scala supports both Highly Scalable and Highly Performance applications?

Answer)As Scala supports Multi-Paradigm Programming(Both OOP and FP) and uses Actor Concurrency Model, we can develop very highly Scalable and high-performance applications very easily

www.smartdatacamp.com

91)What are the available Build Tools to develop Play and Scala based Applications?

Answer)The following three are most popular available Build Tools to develop Play and Scala Applications:

SBT

Maven

Gradle

92)What is Either in Scala?

Answer)In Scala, either is an abstract class. It is used to represent one value of two possible types. It takes two type parameters: Either[A,B].

93)What are Left and Right in Scala? Explain Either/Left/Right Design Pattern in Scala?

Answer)It exactly has two subtypes: Left and Right. If Either[A,B] represents an instance A that means it is Left. If it represents an instance B that means it is Right. This is known as Either/Left/Right Design Pattern in Scala.

94)How many public class files are possible to define in Scala source file?

Answer)In Java, we can define at-most one public class/interface in a Source file.

Unlike Java, Scala supports multiple public classes in the same source file.

We can define any number of public classes/interfaces/traits in a Scala Source file

95)What is Nothing in Scala?

Answer)In Scala, nothing is a Type (final class). It is defined at the bottom of the Scala Type System that means it is a subtype of anything in Scala. There are no instances of Nothing.

96)Whats the difference between the following terms and types in Scala: Nil, Null, None, and Nothing in Scala?

Answer)Even though they look similar, there are some subtle differences between them, let's see them one by one:

Nil represents the end of a List.

Null denotes the absence of value but in Scala, more precisely, Null is a type that represents the absence of type information for complex types that are inherited from AnyRef.

It is different than null in Java. None is the value of an Option if it has no value in it. Nothing is the bottom type of the entire Scala type system, incorporating all types under AnyVal and AnyRef. Nothing is commonly used as a return type from a method that does not terminate normally and throws an exception.

97)How to you create Singleton classes in Scala?

Answer)Scala introduces a new object keyword, which is used to represent Singleton classes. These are the class with just one instance and their method can be thought of as like Java's static methods. Here is a Singleton class in Scala:

```
package test

object Singleton{

def sum(l: List[Int]): Int = l.sum

}
```

This sum method is available globally, and can be referred to, or imported, as the test. Singleton.sum. A singleton object in Scala can also extend classes and traits.

98)What is Option and how is it used in Scala?

Answer)The Option in Scala is like Optional of Java 8. It is a wrapper type that avoids the occurrence of a NullPointerException in your code by giving you default value in case object is null.

When you call get() from Option it can return a default value if the value is null.

More importantly, Option provides the ability to differentiate within the type system those values that can be nulled and those that cannot be nulled.

99)What is the difference between a call-by-value and call-by-name parameter?

Answer)The main difference between a call-by-value and a call-by-name parameter is that the former is computed before calling the function, and the later is evaluated when accessed.

100)What is default access modifier in Scala? Does Scala have public keyword?

www.smartdatacamp.com

Answer)In Scala, if we dont mention any access modifier to a method, function, trait, object or class, the default access modifier is “public”. Even for Fields also, “public” is the default access modifier. Because of this default feature, Scala does not have “public” keyword.

101)Is Scala an Expression-Based Language or Statement-Based Language?

Answer)In Scala, everything is a value. All Expressions or Statements evaluates to a Value. We can assign Expression, Function, Closure, Object etc. to a Variable. So, Scala is an Expression-Oriented Language.

102). Is Java an Expression-Based Language or Statement-Based Language?

Answer)In Java, Statements are not Expressions or Values. We cannot assign them to a Variable. So, Java is not an Expression-Oriented Language. It is a Statement-Based Language.

103)Mention Some keywords which are used by Java and not required in Scala?

Answer)Java uses the following keywords extensively: ‘public’ keyword - to define classes, interfaces, variables etc. ‘static’ keyword - to define static members

104)Why Scala does not require them?

Answer)Scala does not require these two keywords. Scala does not have ‘public’ and ‘static’ keywords.

In Scala, default access modifier is ‘public’ for classes, traits, methods/functions, fields etc.

That’s why, ‘public’ keyword is not required.

To support OOP principles, Scala team has avoided ‘static’ keyword. That’s why Scala is a Pure-OOP Language.

It is very tough to deal static members in Concurrency applications.